Title
Multiscale Geographically Weighted Discriminant Analysis

Authors
Comber, Alexis
Malleson, Nick
Nguyen Thi Thuy, Hang
et al.

Publication Date
2021-09-01

DOI
10.25436/E2PP4F

Peer reviewed

Multiscale Geographically Weighted Discriminant Analysis

Alexis Comber[1]
School of Geography, University of Leeds, UK
a.comber@leeds.ac.uk

Nick Malleson
School of Geography, University of Leeds, UK
n.s.malleson@leeds.ac.uk

Hang Nguyen Thi Thuy
VNU Vietnam Japan University, Hanoi, Vietnam
nguyen.t.thuyhang@gmail.com

Thanh Bui Quang
Faculty of Geography, VNU University of Science, Hanoi, Vietnam
qthanh.bui@gmail.com

Minh Kieu
Faculty of Engineering, University of Auckland, New Zealand
minh.kieu@auckland.ac.nz

Phe Hoang Huu
R&D Consultants, Hanoi City, Vietnam
hoanghuuphe@gmail.com

Paul Harris
Sustainable Agriculture Sciences,Rothamsted Research, UK
paul.harris@rothamsted.ac.uk

Abstract

This paper describes the novel development and application of a multi-scale geographically weighted discriminant analysis (MSGWDA). This is applied to a case study of survey data of attitudes to a proposed motorbike / scooter ban in Han Noi, Vietnam. It uses discriminant analysis to examine attitudes to the ban in relation to travel purposes, distances, respondent age and so on. The main part of the paper focuses on describing the novel MSGWDA approach, and the results indicate the varying scales of relationship between the different input variables and the categorical responses variable. The paper also reflects on the pervasive logic of the approaches used to fit multiscale geographically weighted bandwidths (for example in regression). These have historically been based on the iterative back-fitting approaches used in GAMs, but risk missing potentially important variable interactions amongst un-evaluated bandwidths because of the sequence of their application. It is argued that although pragmatic in the 1990s, it may be possible to apply more deterministic approaches with increased memory and readily accessible computing power in order to better navigate such highly dimensional search spaces.

Funding This work has received funding from the British Academy under the Urban Infrastructures of Well-Being programme [grant number UWB190190]

[1] Corresponding author

1 Introduction

Discriminant analysis (DA) [5, 12], is a commonly used technique for predicting membership or class for discrete groups as an alternative to multinomial logistic regression [10]. Recently DA has gained much attention in the context of machine learning [9] and real time analyses [13] because it can also be used as an information learning technique such as pattern recognition.

Conceptually, in DA the data used as input can be thought of as having been drawn from different populations of each class [2]. The discriminant functions are extracted and then used to generate class membership probabilities for each observation. If there are k groups, the aim is to extract k, under the assumption that the data are multivariate normal, then if $\sum_j$ is the variance-covariance matrix for the members of class j, q is the number of predictor variables in $\mathbf{x}$, μ_j is the mean vector for the observations in class j, and p_j is the prior membership probability of class j, the linear assignment can be written as:

$$k = \arg\ \max_{j\in(1,\ldots,m)} \left[p_j \frac{1}{(2\pi|\Sigma_j|)^{q\backslash 2}} exp\left(-\frac{1}{2}(\mathbf{x}-\mu_j)' {\sum_j}^{-1} \mathbf{x}-\mu_j) \right) \right] \tag{1}$$

LDA was extended from the linear to the quadratic case by Marks and Dunn [11]. DA was further extended to the spatial case by [2] who proposed a geographically weighted DA (GWDA). Whereas a standard DA (LDA and QDA) uses the mean vector and covariance matrix, a GWDA uses geographically weighted means and covariances as described in Brunsdon et al [1] and Fotheringham et al [6]. It uses the same geographically weighted (GW) framework as GWR, in which a series of local models are constructed rather than one global model. However, thinking around GW frameworks has matured considerably in recent years. Multiscale GWR (MSGWR) seeks to identify variable specific bandwidths rather than using a single best on average bandwidth to construct local models. The idea is that individual response-to-predictor relationships may operate over different spatial scales and the use of a single bandwidth in a standard GWR may under- or over-estimate those. As a result MSGWR has been suggested as the default GWR approach [4]. Such thinking and logic has potential relevance for all GW frameworks, including GWDA, hence the method proposed in this paper

2 Multiscale Geographically Weighted Discriminant Analysis

In GWDA the population probabilities depend on the spatial location of the observation – ie the variance-covariance matrix $\sum_j$, the prior membership probabilities of class j, p_j or the μ_j the mean vector for the observations in class j, are assumed to vary with spatial location $\mathbf{u}$. Thus, the probabilities used to derive the decision rules are conditional on $\mathbf{u}$:

$$f_p(\mathbf{x}|\mathbf{u}) = \frac{1}{(2\pi|\Sigma_j(\mathbf{u})|)^{q/2}} exp\left(-\frac{1}{2}(\mathbf{x}-\mu_j(\mathbf{u}))' {\sum_j}^{-1}(\mathbf{u})(\mathbf{x}-\mu_j(\mathbf{u})) \right) \tag{2}$$

The key objective in all multiscale GW models is to determine the matrix of parameter specific weights. These in this case will be used to weight each input variable at location $\mathbf{u}$, as defined by the kernel bandwidth. Figure 1 shows an example of the different bandwidths and potential scales of relationship between the classification and different variables.

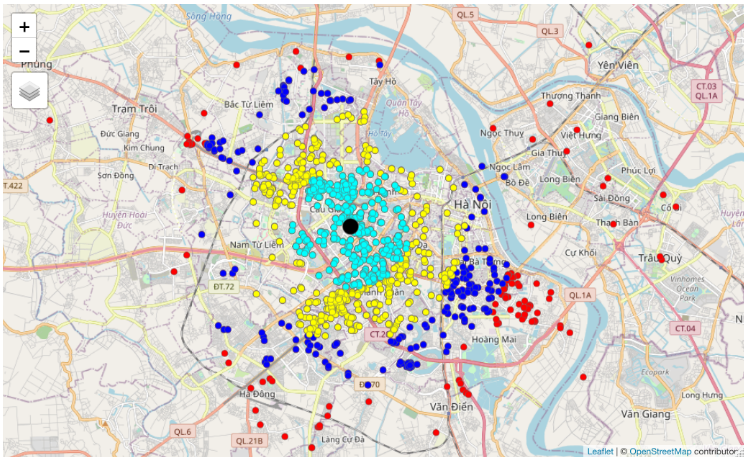

Figure 1 An illustration of the different adaptive bandwidths, shaded in cyan (30%), yellow (70%), blue (90%) and red (100%), for 4 different variables, for a location marked in black, with an OpenStreetMap backdrop.

3 Case study: Travel Survey in Ha Noi

Ha Noi like many major cities in emerging economies, suffers serious traffic congestion and air pollution due to rapid urbanization rates, increases in private transport. Motorbikes are the preferred transportation mode: almost everyone in the city owns a motorbike. In 2015, Ha Noi had 4.9 million motorbikes and 11 million motorbikes are projected by 2025. As a result the government in Vietnam is exploring the possibility of implementing a motorbike ban. A survey has been undertaken to capture attitudes to the ban as part of an ongoing project and was thus used for this study. Data from 1191 respondents was obtained and used in the analyses as described below. The aim was to examine create a MSGWDA of attitudes to the ban from categorical variables describing:

- respondent age group;
- respondent gender;
- the purpose of the main regular journey they make;
- the network distance of that journey, as derived from a shortest path analysis of OSM route data with snap distances.

To demonstrate MSGWDA, combinations of adaptive bandwidth sizes for each variable were defined as sequences running from 20% to 100% in steps of 10%. For 4 variables, this resulted in 9^4 bandwidth combinations to evaluate. Each combination of variable specific bandwidths was used to weight inputs into a linear discriminant analysis function (`lda` part of the `MASS` R package). For simplicity a boxcar weighting was used, generating weights of 1 for observations underneath the kernel and 0 for those outside. These were used to create a locally weighted LDA at each observation location which was used to make a local ban attitude prediction. The entire set of predictions were then evaluated using overall and Kappa accuracies. The best performing combinations of bandwidths was then identified.

4 Results

Two results are used to illustrate the potential inferential advantages of the MSGWDA: an ordinary global LDA and a novel multiscale GWDA. The standard LDA model is relatively

Figure 2 The motorbike ban attitudes of the survey respondents, with a density surface of respondent home locations (band = 0.01 degree; bins = 16), and a Stamen toner backdrop.

weak, with an overall accuracy of 0.548 and a Kappa statistic of 0.115. The correspondence table is shown in Table 1 and indicates high specificity (ie good at true negatives) and low sensitivity (ie poor at true positives).

	Observed		
Predicted	agree	disagree	neutral
agree	42	24	24
disagree	222	585	238
neutral	14	16	26

Table 1 The correspondence matrix of the LDA of survey responses.

The MSGWDA examined combinations of adaptive bandwidths for each variable. For each of these, a geographically weighted LDA model was created at each of the 1191 respondent home locations. At each location a weighted LDA model was used to predict the motorbike ban attitude, such that a vector of 1191 predicted ban attitudes were created from 1191 local models. For each set of predictions, a correspondence matrix of predicted against observed ban attitudes was created an evaluated using overall accuracy and Kappa statistics. The best performing combinations were found to be the following sets of bandwidths when evaluated using Overall accuracy and Kappa statistics:

- Overall accuracy: gender 80%, trip purpose 50%, age 40% and network distance 10%.
- Kappa statistic: gender 40%, trip purpose 20%, age 20% and network distance 10%.

These are illustrated in Figure 3 for the same example location as in Figure 2. Here we can see the different bandwidths indicated by different fit or accuracy measures. The correspondences are summarised in Table 2 and result in Overall accuracies and Kappa statistics of 0.579, 0.199 and 0.575, 0.207, respectively.

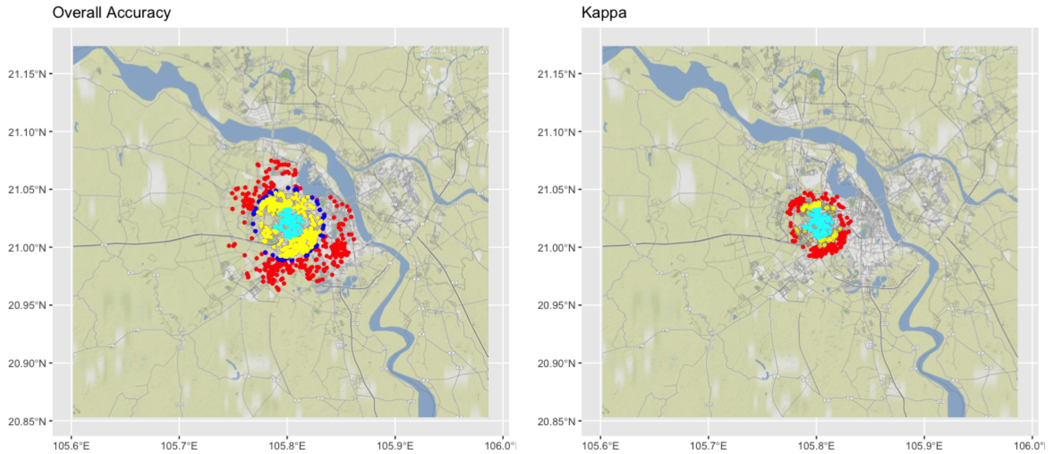

Figure 3 An illustration of the best multiscale bandwidths, evaluated using Overall Accuracy and Kappa statistic (Gender in red, Trip purpose in blue, Age in yellow and Distance in cyan).

	Overall			**Kappa**		
Predicted	agree	disagree	neutral	agree	disagree	neutral
agree	59	27	22	65	37	30
disagree	199	573	209	191	556	194
neutral	20	25	57	22	32	64

Table 2 The correspondence matrices of the MSDWDA classifications of survey responses, when evaluated using Overall accuracy and Kappa statistics.

5 Discussion

The MSGWDA approach improves the classification accuracy compared to a standard global LDA but importantly also indicates the variations in the spatial scales at which categorical data are associated with the outcome: the gender variable tends towards the global, with trip purpose, age and distance highly localised in their effect. This understanding of scale will inform future project work in relation to the transport and behaviour simulation models being developed within this project.

Arguably the major discussion point to arise from this work has been due to the need to unpick the mechanisms of multiscale GW models. The key question arising from the back-fitting methods they employ is this:

> How confident can we be that that potentially important variable interactions are not being missed by this *fix the first variable bandwidth, then fix the second, then the next, etc, etc ...* approach, rather than looking at all possible combinations of bandwidths?

The answer to this is uncertain: the multivariate bandwidth search space to determine the optimal set of weights to be passed to the local model at location on **u** is potentially huge. In the past, pragmatic short-cuts were needed to be able to move through it. But times and computing power have both changed. The original MSGWR [14, 7] and subsequent refinements were based on the approach taken in generalized additive models (GAMs) [8, 3]. Essentially what these do to determine the optimal set of bandwidths is to determine the bandwidth for each variable sequentially, using smoothing functions that assume the other terms are known. We suspect that this approach was developed by the GAM team as a pragmatic way overcoming the difficulty in searching through a high dimension solution space comprised of all possible bandwidths for all possible variables. It was then adopted by the initial work into MSGWR due to the high dimensionality of the solution search space (2000 observations with 5 explanatory requires $2000^6 = 6.4 \times 10^{19}$ solutions to be evaluated for a regression (including the intercept) and $2000^5 = 3.2 \times 10^{16}$ for a discriminant analysis. With potentially greater computing power a grid of all possible combinations of parameter specific bandwidths could be evaluated. This is philosophically preferable: the specification of multiscale bandwidths one parameter at a time potentially ignores variable interactions at scales not considered in previously fixed bandwidths. Future work will definitely explore this in greater detail!

References

1. Chris Brunsdon, AS Fotheringham, and Martin Charlton. Geographically weighted summary statistics—a framework for localised exploratory data analysis. *Computers, Environment and Urban Systems*, 26(6):501–524, 2002.
2. Chris Brunsdon, Stewart Fotheringham, and Martin Charlton. Geographically weighted discriminant analysis. *Geographical Analysis*, 39(4):376–396, 2007.
3. Andreas Buja, Trevor Hastie, and Robert Tibshirani. Linear smoothers and additive models. *The Annals of Statistics*, pages 453–510, 1989.
4. Alexis Comber, Chris Brunsdon, Martin Charlton, Guanpeng Dong, Rich Harris, Binbin Lu, Yihe Lü, Daisuke Murakami, Tomoki Nakaya, Yunqiang Wang, et al. The gwr route map: a guide to the informed application of geographically weighted regression. *arXiv preprint arXiv:2004.06070*, 2020.
5. Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2):179–188, 1936.
6. A Stewart Fotheringham, Chris Brunsdon, and Martin Charlton. *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons, 2003.
7. A Stewart Fotheringham, Wenbai Yang, and Wei Kang. Multiscale geographically weighted regression (mgwr). *Annals of the American Association of Geographers*, 107(6):1247–1265, 2017.
8. Trevor Hastie and Robert Tibshirani. Generalized additive models: some applications. *Journal of the American Statistical Association*, 82(398):371–386, 1987.
9. Tyler L Hayes and Christopher Kanan. Lifelong machine learning with deep streaming linear discriminant analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 220–221, 2020.
10. David W Hosmer and Stanley Lemeshow. Confidence interval estimates of an index of quality performance based on logistic regression models. *Statistics in medicine*, 14(19):2161–2172, 1995.
11. Sidney Marks and Olive Jean Dunn. Discriminant functions when covariance matrices are unequal. *Journal of the American Statistical Association*, 69(346):555–559, 1974.
12. C Radhakrishna Rao. The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society: Series B (Methodological)*, 10(2):159–193, 1948.
13. Daniel Szostak, Krzysztof Walkowiak, and Adam Włodarczyk. Short-term traffic forecasting in optical network using linear discriminant analysis machine learning classifier. In *2020 22nd International Conference on Transparent Optical Networks (ICTON)*, pages 1–4. IEEE, 2020.
14. Wenbai Yang. *An extension of geographically weighted regression with flexible bandwidths*. PhD thesis, University of St Andrews, 2014.