

RANDOM FORESTS TO EVALUATE ANTIMICROBIAL MULTIRESTANCE IN SALMONELLA TYPHIMURIUM

Marzia Mancin, Marco Ruffa, Lisa Barco, Antonia Anna Lettini, Antonia Ricci
Public Health and Risk Analysis Department; Istituto Zooprofilattico delle Venezie

INTRODUCTION

Antimicrobial resistance in zoonotic bacteria is a problem of increasing concern worldwide and multiresistance is a critical aspect of such phenomenon, since it can lead to human infections untreatable with available drugs. In order to counteract the diffusion of resistance, scientists have to provide policy makers with relevant data on factors contributing to its emergence and spreading, and statistical tools like Random Forests (RF) are used to guarantee the significance of results and to facilitate their interpretation.

MATERIALS AND METHODS

- > A selection of 556 *S. Typhimurium* strains isolated in Italy from different types of sample (animals, foodstuffs and environment) during the period 2004-2009 was submitted to antimicrobial resistance analysis by disk diffusion;
- > A panel of 15 antibiotics was used to evaluate the antimicrobial multiresistance (AMR), defined as resistance to more than 3 antimicrobials;
- > RF was used to verify which variables among years, districts, material and place of sampling, species, type of sample and phagetype are more important to explain the AMR in terms of mean decrease accuracy (MDA) and mean decrease of Gini (MDG)
- > The AMR was considered as:
 - binomial variable: 0=no multiresistant, 1=resistant to more than 3 antimicrobials; i.e. 2 way classification;
 - multinomial variable: 0=no multiresistant, 2=resistant to more than 3 antimicrobials, 1=otherwise; i.e. 3 way classification; resistant to K=0, 1...6 antimicrobial; i.e. K way classification.

RESULTS

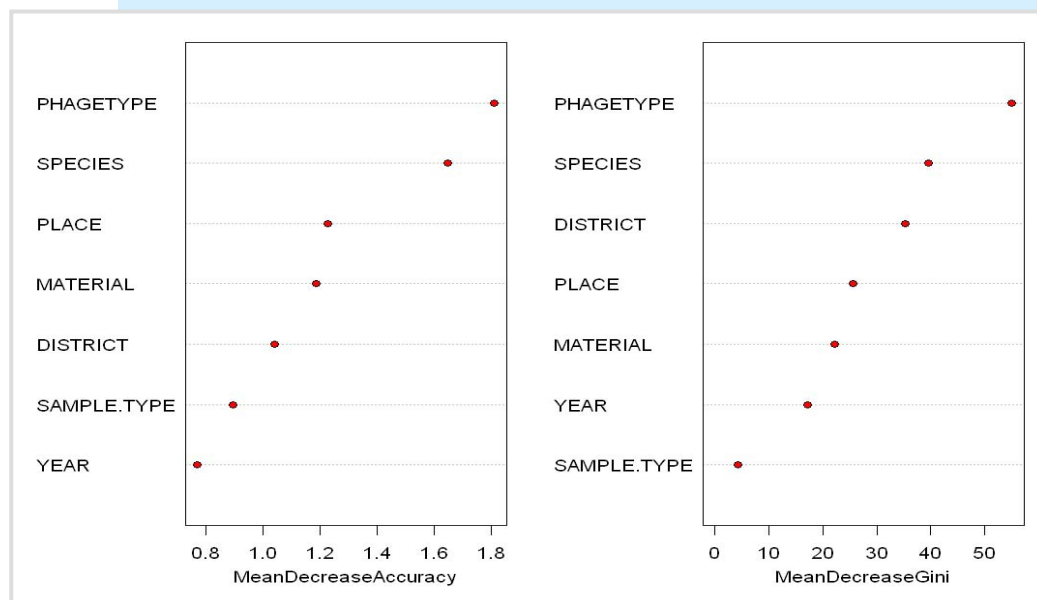


Figure a: Classification of covariates for the AMR in accordance with MDA and MDG.

RF output shows that phagetype and species are two covariates that explain most of the AMR, whereas year and type of sample are not important (Fig. a). To evaluate the accuracy of classification method it is possible to design Multi-dimensional Scaling Plot to visualize the dissimilarity data in a two dimensional space and the convergence of error that shows the errata classification of model per category (Fig. b). In the case of binomial AMR it is also possible to design the ROC curve (Fig. c). The increase of AMR classification way produces a rise in average rate of error classification (Tab. 1).

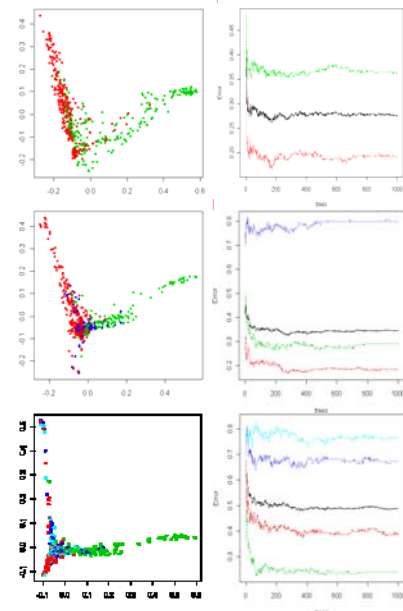


Figure b: Multi-dimensional Scaling Plot and Convergence of error per category in 2-way, 3-way and 4-way analysis.

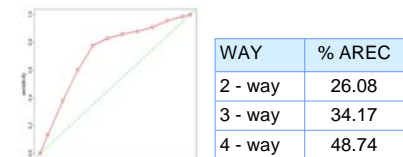


Figure c: 2-way ROC curve.

WAY	% AREC
2 - way	26.08
3 - way	34.17
4 - way	48.74

Table 1: Average rate of Error Classification of Random Forest.

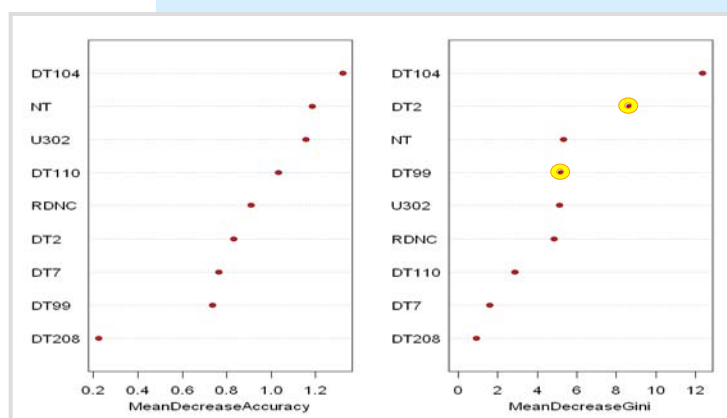


Figure d: Classification of covariates for the AMR in accordance with MDA and MDG.



Figure e: Classification tree of phagetypes.

Phag.	LDA	LR	p-value
DT104	2.04	1.82	0.000
NT	1.04	0.82	0.000
U302	1.92	1.68	0.001
DT110	2.47	2.53	0.017
DT7	1.70	1.43	0.037
RDNC	-0.35	-0.29	NS
DT2	-2.13	-14.86	NS
DT99	-2.13	-14.87	NS
DT208	-0.74	-0.64	NS

Table 2: Coefficients of LDA and LR regression.

RF applied to different phagetypes shows that DT104, NT, U302, DT110 (multiresistant) are the most significant phagetypes for the prediction accuracy of model (MDA) of AMR; MDG indicates that, besides these variables, also DT2 and DT99 (not multiresistant) are important for the splitting of dependent variable (Fig. d).

RF is consistent with the results of the classification tree (Fig. e), of the Linear Discriminant Analysis (LDA) and of the Logistic Regression (LR) (Tab. 2) or Multinomial Regression (MR) (2 or more way respectively). The classification error for these different methods is in this case very similar.

CONCLUSIONS

RF yield manageable results, allow to analyze high amounts of variables and to assure non parametric robust results. RF represents a valid alternative to other methods of multivariate analysis such as LDA and MR.