Geo.info

40

Extended Abstract Proceedings of the

GIScience 2014

Compiled and Edited by
Kathleen Stewart
Edzer Pebesma
Gerhard Navratil
Paolo Fogliaroni
Matt Duckham

Department of Geodesy and Geoinformation
Vienna University of Technology

Geo.info

40

Extended Abstract Proceedings of the GIScience 2014

Compiled and Edited by by
Kathleen Stewart
Edzer Pebesma
Gerhard Navratil
Paolo Fogliaroni
Matt Duckham

Department of Geodesy and Geoinformation
Vienna University of Technology
Gusshausstraße 27-29/120.2
1040 Vienna, Austria

Series Editor

Andrew U. Frank
Dept. of Geodesy and Geoinformation
Vienna University of Technology
Gusshausstr. 27-29/120.2
A-1040 Vienna, Austria
frank@geoinfo.tuwien.ac.at

ISBN 978-3-901716-42-3 GeoInfo Series Vienna

Printed in Austria

Typesetting: Camera ready by author/editor
Printing and binding: Hochschülerschaft, TU Vienna
Wirtschaftsbetriebe GmbH

Sensitivity analysis of Support Vector Machine land use change modelling method

Mileva Samardžić-Petrović[1], Branislav Bajat[1], Miloš Kovačević[1], Suzana Dragićević[2]

[1] Faculty of Civil Engineering, University of Belgrade, Bulevar kralja Aleksandra 73, 11000 Belgrade, Serbia
Email: {mimas; bajat; milos}@grf.bg.ac.rs

[2] Spatial Analysis and Modeling Laboratory, Department of Geography, Simon Fraser University, 8888 University Dr., Burnaby, B.C., V5A1S6, Canada
Email: suzanad@sfu.ca

1. Introduction

When using machine learning (ML) approaches for land use (LU) change modelling the main goal is to find a function that is the best approximation of the nonlinear problem that represents the complex LU change process. Support Vector Machines (SVM) is one ML method capable of solving nonlinear problems and has been applied to various disciplines such as ecology (Drake et al. 2006), hydrology (Tripathi et al. 2006) and remote sensing (Brown et al. 2000). Interest in using the SVM method for LU changes modelling has grown in recent years (Yang et al. 2008, Okwuashi et al. 2012). However, as a relatively new method, SVM is insufficiently researched in LU change modeling, particularly in relation to its sensitivity to parameter changes, attribute selection, data sampling choices, and data representation.

The main objective of this research study is to conduct a sensitivity analysis investigation on a SVM-based LU change model with respect to attribute selection and parameter changes. The efficient application of ML methods, including SVM, requires the selection of appropriate attributes (features). Attribute selection is an important stage of modelling as some attributes can have small or no predictive power at all, and hence "confuse" the ML process. Various methods for attribute selection exist. However, this study uses Info Gain (IG), Gain Ratio (GR) and Correlation-based Feature Subset (CFS) (Witten et al. 2011). Moreover, the efficient application of SVM requires the selection of an optimal combination of parameters. Because the Radial Basis Function (RBF) (Abe, 2005) was used as the kernel function for SVM, model sensitivity was analyzed for changes to two parameters; γ of the RBF and penalty C. Method have

2. Experiment

2.1 Study area and data representation

The study area included the Zemun Municipality, part of the territory of the City of Belgrade, Republic of Serbia. In 2013, the administrative area was approximately 17.5km x 8km and included the old city and two suburban areas.

The data used for this study included: three orthophoto images for the years 2003, 2007 and 2011; maps of actual LU classes obtained from the Urban Planning Institute of Belgrade, and; publicly available population census data (Statistical Office of the Republic of Serbia).

The study area was represented as a rectangular cell grid of 10m spatial resolution. Also, each grid cell was represented as an n-dimensional real vector $\mathbf{x}^t$ ($\mathbf{x}^t = <x^t_1, x^t_2, \ldots, x^t_n>$), where coordinate x^t_i represents the value of the i-th attribute (corresponding to LU class and created urban attributes) associated with the cell x, at a particular time t. The various urban attributes considered were: Euclidian distance ed to municipality centre, city centre (old core of Belgrade), Danube and Sava rivers, green areas, railway, highway, main road, streets of category I and II, and Population Change Index (PCI) between two censuses for the years 2002 and 2011. PCI provides a standardized measure for comparing population changes over time (Bajat et al. 2013).

Since the goal of the SVM model is to predict future land use changes, it is necessary to prepare a minimum of two datasets to represent the study area at three different moments in time (t-1, t and t+1). In order to build the model $\mathbf{x}^{t-1} \rightarrow y^t$, SVM uses training dataset ($\mathbf{x}^{t-1}$, y^t), which contains $\mathbf{x}^{t-1} = \langle x^{t-1}_1, x^{t-1}_2, \ldots, x^{t-1}_n \rangle$ as input attributes and y^t as the output attribute to be predicted. For this study, y^t contains nine LU classes: agricultural, wetlands, traffic areas, infrastructure, residential, commercial, industry, special use and green areas.

Based on the developed SVM model, $\mathbf{x}^{t-1} \rightarrow y^t$, and by using $\mathbf{x}^t = \langle x^t_1, x^t_2, \ldots, x^t_n \rangle$ as input attributes, y^{t+1}_p can be predicted. Therefore, the second dataset is a test dataset ($\mathbf{x}^t$, y^{t+1}) and it is used for independent validation of the built SVM model, achieved by comparing the predicted (y^{t+1}_p) and real (y^{t+1}) LU classes at time t+1. The training dataset was created based on data from years 2003 and 2007, and the test data created from years 2007 and 2011.

The Kappa statistics was used to compare the model output with the real land use map for year 2011. The overall land use change for the study time period (2003-2011) was small (3%) relative to the overall study area size. Hence, in order to obtain more informative datasets it was necessary to conduct data sampling. Consequently, in this study the training and test datasets were created to contain the same number of changed and unchanged cells, thereby ensuring that all LU classes are proportionally represented.

2.2 Attribute Selection

From datasets S containing all of the created attributes, three subsets of attributes were selected using the IG, GR and CFS methods, and the results shown in Figure 1. The IG and GR methods rank the attributes independently of each other based on their measure of association with the LU class in time t, while the CFS method automatically determines a subset of k relevant attributes that are highly correlated with the LU class but uncorrelated with each other.

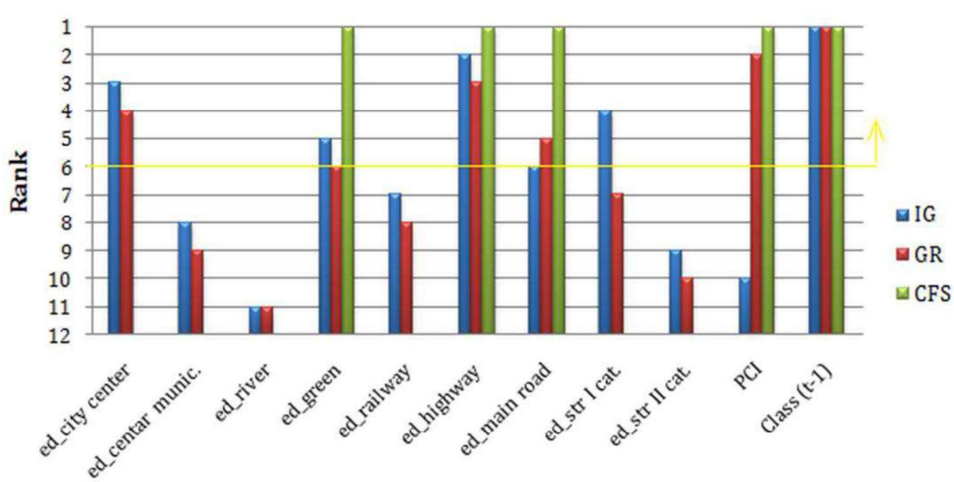

Figure 1. Comparisons of the attribute selection by IG, GR and CFS methods.

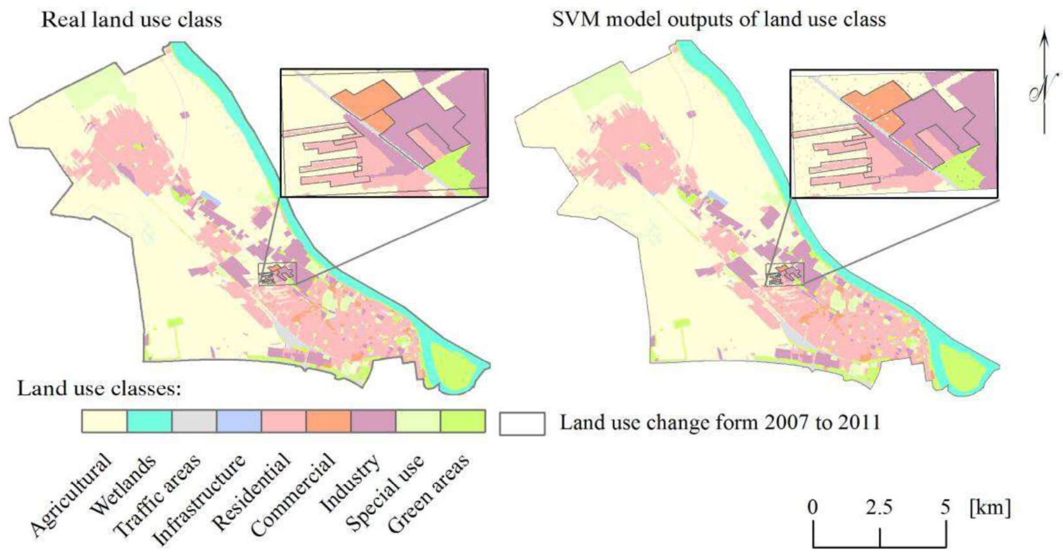

Figure 2. Generated land use changes with SVM model for 2011.

One of the obtained SVM model outputs for the generated land uses for year 2011 is shown in Figure 2. Moreover, the results indicate the CFS method selected subsets of five attributes. Therefore, in order to compare the sensitivity of models built with attributes selected using the three methods with respect to the SVM parameters, the five highest ranked attributes by IG and GR were selected and three datasets were created. Each dataset S^{CFS}, S^{IG} and S^{GR} contains training and test datasets for five selected attributes respectively for the CFS, IG and GR methods.

2.2 Sensitivity of model on SVM parameters and selected attributes

In order to demonstrate the significance of the attribute selection, additional models were created based on the datasets S. Values of validation measure for all models, built by using various SVM parameters and four different data representations are shown in Figure 3.

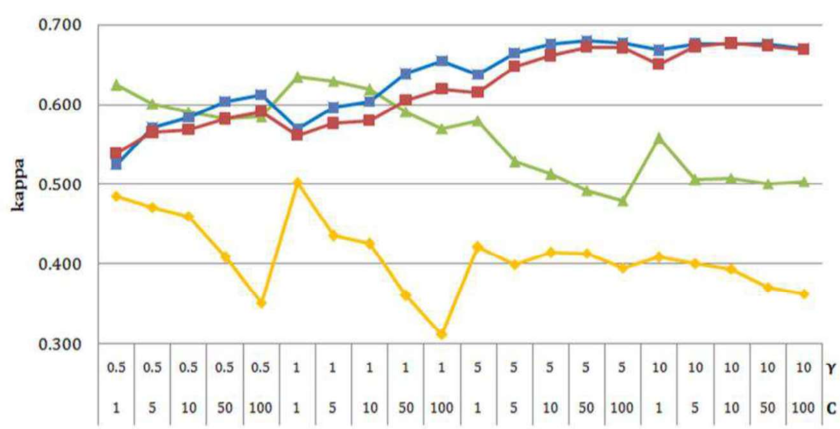

Figure 3. Kappa values for LU change models based on different values for SVM parameters γ and C for S, S^{CFS}, S^{IG} and S^{GR} datasets.

The results indicate that the S^{CFS} and S^{GR} datasets exhibit better kappa performance and are more robust to different SVM parameter combinations. Models built based on S^{CFS} datasets are slightly better than the ones based on S^{GR}. Using S and S^{IG}, the LU model have less capability to predict changes and can be overfitted with higher values of parameters.

3. Conclusion

The obtained results indicate the subset of the same number of attributes selected by the CFS and GR methods increased kappa values, while attributes selected by the IG method decreased kappa values comparing to models built using all attributes. Using selected attributes by the CFS and GR methods resulted in a simple (less attributes – less complicated) model but with better performance and with less possibility to be overfitted with higher values of parameters. For the datasets used, the subset of k attributes selected by the CFS method provided slightly better models compared to the k highest ranked attributes by GR, and significantly better models compared to k highest ranked attributes by IG. In order to further explore sensitivity analysis of the SVM modelling approach, the future work will include using datasets covering different study areas and with different land use change dynamics. Furthermore, the k number of selected attributes was not investigated as the optimal number of attributes selected by IG and GR, which will also be pursued in future research.

Acknowledgements

This study was supported by an award from the Ministry of Education, Science and Technological Development of the Republic of Serbia (Project No. III 47014) to the first three authors, and by a Natural Sciences and Engineering Research Council (NSERC) of Canada Discovery Grant awarded to the fourth author.

References

Abe S, 2010, Support vector machines for pattern classification, 2nd ed: Springer.

Bajat B, Krunić N, Samardzić-Petrović M and Kilibarda M, 2013, Dasymetric modelling of population dynamics in urban areas, Geodetski vestnik, 2013(57): 777–792.

Brown M, Lewis HG and Gunn SR, 2000, Linear spectral mixture models and support vector machines for remote sensing, Ieee Transactions on Geoscience and Remote Sensing 38 (5):2346-2360.

Drake JM, Randin C and Guisan A, 2006, Modelling ecological niches with support vector machines, Journal of Applied Ecology 43 (3):424-432.

Okwuashi O, McConchie J, Nwilo P, Isong M, Eyoh A, Nwanekezie O, Eyo E and Ekpo AD, 2012, Predicting future land use change using support vector machine based GIS cellular automata: a case of Lagos, Nigeria. Journal of Sustainable Development 5 (5):132-139.

Tripathi S, Srinivas VV and Nanjundiah R S, 2006, Dowinscaling of precipitation for climate change scenarios: A support vector machine approach, Journal of Hydrology 330 (3-4):621-640.

Witten IH, Frank E, and Hall MA, 2011, Data Mining: Practical Machine Learning Tools and Techniques: Practical Machine Learning Tools and Techniques. 3rd ed: Elsevier.

Yang Q, Li X, and Shi X, 2008, Cellular automata for simulating land use changes based on support vector machines, Computers & Geosciences 34 (6):592-602.