

Supervised learning for automated feature selection in road network generalization

Izabela Karsznia ^{a,*}, Albert Adolf ^a, Stefan Leyk ^b

^a Department of Geoinformatics, Cartography and Remote Sensing, Faculty of Geography and Regional Studies, University of Warsaw, Poland; i.karsznia@uw.edu.pl, a.adolf@student.uw.edu.pl

^b Department of Geography, University of Colorado Boulder; stefan.leyk@colorado.edu

* Corresponding author

Keywords: road network, cartographic generalization, selective omission, small-scale maps, machine learning, data enrichment

Abstract:

Selection constitutes the first step in cartographic generalization, often in form of a vector generalization operator that influences the visual quantity of the map by reducing its content or visual clutter (Stanislawski et al. 2014). Selection is usually followed by operators dealing with aspects of visual quality and aesthetics such as road simplification, smoothing or displacement. Over the years, road network generalization has gained increasing attention, and while various methods for road selection have been proposed, there is still a lack of effective solutions dedicated to small scale maps (Benz and Weibel 2014). Existing approaches are often not effective enough to be implemented for map production. To improve the performance of such selection approaches prior research has suggested the use of additional information contained in other sources but the implementation of such measures such as local road density variations remains a computational challenge (Stanislawski et al. 2014). Recently, approaches based on machine learning have been shown as promising solution for settlement selection at small scales (Karsznia and Weibel 2018; Karsznia and Sielicka 2020), the generalization of buildings (Sester et al. 2018, Feng et al. 2019, Li et al. 2020) as well as the smoothing and selecting of line objects (Zhou and Li 2014; Karsznia et al. 2020), including road features in mountainous settings (Courtial et al. 2020).

This research aimed to automate a selective omission process in road generalization, making the map content as close as possible to that of a manual map design. To achieve this, we used data enrichment and machine learning (ML). The General Geographic Objects Database (GGOD) used as a source in this research lacks relevant information about roads. Thus, we used data enrichment techniques to add important semantic information concerning the roads. We then explored the potential of different ML techniques for automated road selection and compared the results with roads selected based on a set of rules defined by the Polish National Map Agency (NMA) to produce well-designed paper maps at small scales. We tested our road network selection approaches within three districts using the GGOD at the source scale of 1:250,000 to create a generalized map at the target scale of 1:500,000. The three selected districts differ in road density, road network structure, and other road characteristics such as the number of intersections. The considered districts are distant from to cover various regions of Poland.

We examined two selection approaches: basic and enhanced. The basic approach used a set of rules provided by the Polish NMA guidelines (Regulation 2011), limited to very basic road characteristics, namely road class, road management category and number of lanes. The enhanced approach was based on supervised learning algorithms after the GGOD had been enriched with additional road characteristics including: road class, road management category, number of lanes, road surface material, road length, centrality measures (betweenness, closeness, degree), town connectivity, road connectivity (which respectively indicate number of roads connected to settlements and number of road segments connected to other roads) and road length in hexagonal reference fields. Information about road status provided by the National Atlas of Poland was also considered in the enhanced approach to explore how roads were selected or omitted by an experienced cartographer. We digitized the road network from the atlas map within the test districts to create appropriate training data for the supervised learning process. Having gathered holistic road characteristics and the training data, we performed binary classification in RapidMiner Studio 9.1.0. We compared performance across various learning models, including decision tree (DT), decision tree with genetic algorithm (DT-GA), random forest (RF), support vector machines (SVM), and neural net (NN). The results indicate for each model which road segments meet model requirements and should be selected and omitted. Additionally, DT-based models provide weights for the individual variables. The results of both approaches were compared with the atlas map at the target scale. The most important variables for road network selection were road length, road connectivity and closeness (one of the considered centrality measures).

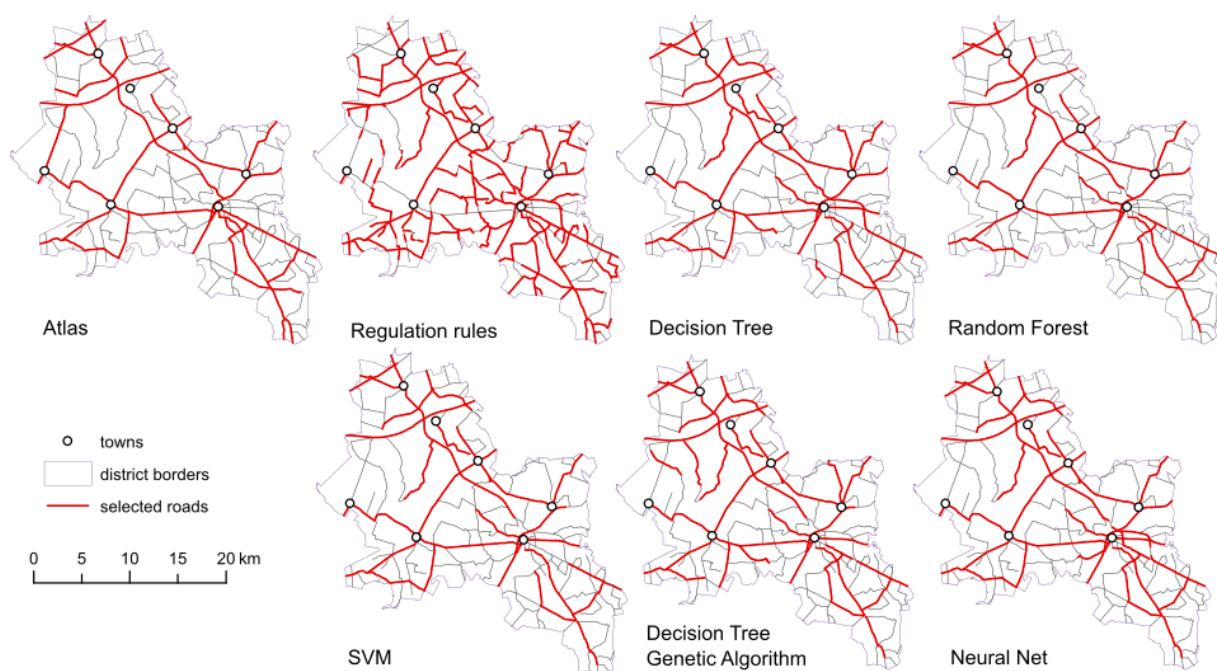


Figure 1. Selection results in the Żyrardowski district at scale 1:500,000.

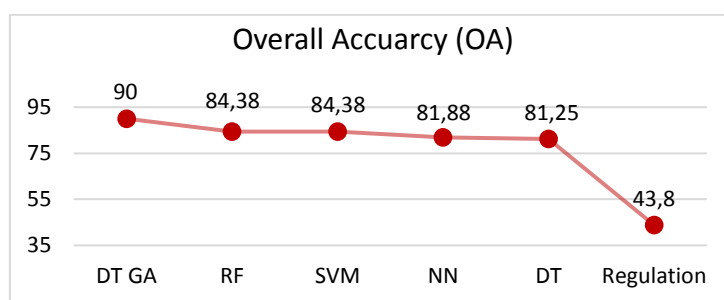


Figure 2. Overall selection accuracy of enhanced (DT GA – DT) and basic (Regulation) approaches for road network generalization to the target scale of 1:500,000.

For each of the five tested ML models we measured the similarity between the achieved results and the reference atlas map (which was manually designed by experienced cartographers). For each approach, basic and enhanced, the similarity was calculated as the number of road segments selected and omitted in comparison to the atlas map. Figure 1 shows the results obtained for the Żyrardowski district. The visual assessment of the results indicated better performance of the enhanced selection approach in three districts: most of the main roads were preserved, and the road network is characterized by a low number of disjointed road segments. In the case of the Żyrardowski district, the most similar selection was obtained by the DT-GA algorithm. While the road network derived using the basic approach (Regulation) was too dense in comparison to the atlas map, some main roads were wrongly omitted, and many selected segments remained discontinuous (disconnected). The selection accuracies for both approaches are presented in Figure 2. The use of ML approaches almost doubled the selection accuracy in comparison to the set of rules used in the Regulation scenario, implemented in the basic approach. This research aimed to explore the possibilities of reducing the cost of map design at different target scales while making it more efficient and faster than manual road network selection. None of the listed models delivered perfect agreement with the atlas map. Thus, we recommend further optimizing steps to be taken in automatic road selection. In future work, we will use additional variables and extend to other test areas to better understand how scalable these approaches are.

Acknowledgements

This research was funded by the National Science Centre, Poland, grant number UMO-2020/37/B/HS4/02605, “Improving Settlement and Road Network Design for Maps of Small Scales Using Artificial Intelligence and Graph Theory”.