International Cartographic Association
Association Cartographique Internationale

Data enrichment and supervised learning in road and settlement feature selection for small scale maps

Albert Adolf [a,*], Karolina Wereszczyńska[a], Izabela Karsznia[a], Robert Weibel[b], Stefan Leyk[c]

[a] *Department of Geoinformatics, Cartography and Remote Sensing, Faculty of Geography and Regional Studies, University of Warsaw, Poland; a.adolf@uw.edu.pl, km.wereszczynska@uw.edu.pl, i.karsznia@uw.edu.pl*

[b] *Department of Geography, University of Zurich, Switzerland; robert.weibel@geo.uzh.ch*

[c] *Department of Geography, University of Colorado Boulder; stefan.leyk@colorado.edu*

* Corresponding author

Keywords: supervised learning, feature selection, roads, settlements, small scales

Abstract:

Selection of appropriate features to be shown on smaller-scale maps constitutes a crucial task in cartographic generalisation. Generally, the selection process is carried out by a vector generalisation operator that modifies the visual quantity of the map by reducing visual clutter (Stanislawski et al. 2014). Thoughtful selection is essential as only the most important and the most characteristic features should be preserved at the target scale. Historically, small scale maps were the result of manual efforts and subjective decisions made by cartographers. Over the years, automated cartographic generalisation has gained increasing attention in the GIScience field, and while various methods for large scale maps have been proposed, there is still a lack of effective solutions dedicated to small scale maps. Existing approaches often underperform because they do not grasp the specifics of small-scale maps. Recently, approaches based on machine learning have shown promise in settlement selection (Karsznia and Weibel 2018; Karsznia and Sielicka 2020), the generalisation of buildings (Sester et al. 2018, Feng et al. 2019, Li et al. 2020) as well as the smoothing and selection of line objects (Zhou and Li 2014; Karsznia et al. 2020), including road features in mountainous areas (Courtial et al.2020).

In this research we aim to extend and evaluate a recently described approach for automatic vector feature selection based on data enrichment and machine learning techniques at small map scales. The main objects of interest are roads and settlements. We apply, refine and evaluate the methods, originally developed for road selection within 6 Polish districts (Karsznia et al. 2021, Karsznia et al 2022), on a much larger data sample. We use 39 additional districts contained in the General Geographic Objects Database at scale 1:250,000 with the goal to select roads at the target scale 1:500,000. We also test and evaluate our recently described settlement selection approach developed for Polish settlements (Karsznia & Weibel 2018; Karsznia & Sielicka 2020) for settlement features in Swiss topographic maps. We examine the approach for three test cantons of Switzerland using settlement data in the TLMRegio database at scale 1:200,000 to create maps at target scale 1:500,000 and 1:1 000 000.

As a first step, the source data for both roads and settlements were enriched with additional semantic, contextual and spatial variables. For road data we added the following variables: road length, road category, road class, surface material, number of lanes, various centrality measures (including road betweenness, closeness, load and degree), number of roads connected to other road segments (road connectivity) and number of settlements proximate to the road segments (settlement proximity). For settlement selection we enriched the database by the following variables: population, urban area, settlement sacral function, settlement sport function, settlement monumental function, commune type, Voronoi area, distance to nearest settlement (nearest neighbour), and settlement density in hexagonal units of 100 km^2 (local settlement density). In the next step two parallel supervised learning processes were tested for roads and settlements selection using the enriched databases and small-scale reference atlas maps as training data for the learning process. We tested and compared decision trees (DT), decision trees supported with genetic algorithms (DTGA), random forest (RF), deep learning based on a multi-layer feed-forward artificial neural network that is trained with stochastic gradient descent using back-propagation (DL), support vector machine (SVM) and back propagation neural network (NN). We used various performance measures including accuracy, precision, recall and F1 measure. For the settlement selection, depending on the scale and across learning models applied, accuracy ranged from 76.35% to 95.01%, precision from 76.35% to 91.30%, recall from 34.10% to 100%, and the F1 measure from 48% to 86.65%. For the road network selection, accuracy ranged from 62.72% to 72.07%, precision from 61.67% to 72.34%, recall from 38.41% to 80.27% and the F1 score from 49.41% to 69.75%. Figures 1 and 2 provide a visual comparison of selected models.

Abstracts of the International Cartographic Association, 6, 3, 2023.
31st International Cartographic Conference (ICC 2023), 13–18 August 2023, Cape Town, South Africa.
https://doi.org/10.5194/ica-abs-6-3-2023 |

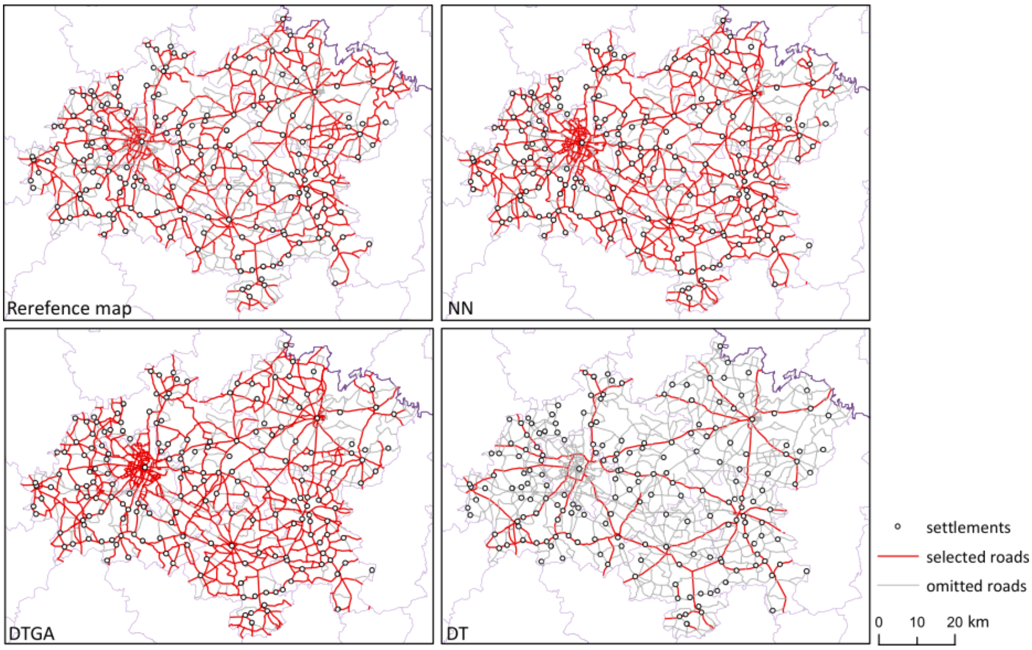

Figure 1. Example results of road network selection in Poznań, Środa Wielkopolska, Gniezno and Września districts (Poland) for a target scale of 1:500,000.

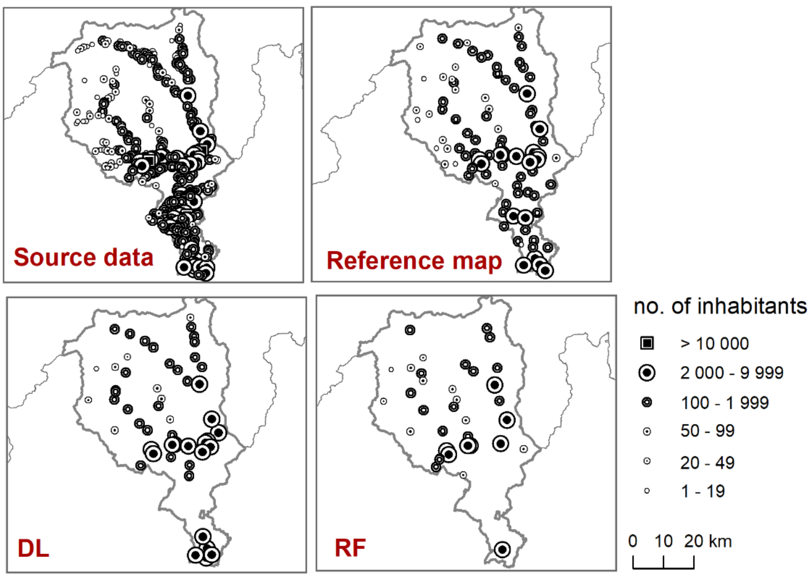

Figure 2. Example results of settlements selection in Ticino (Switzerland) for a target scale of 1:500,000.

In the case of road selection, the most important variables were road length, management category, number of lanes, degree, betweenness, closeness, road class, road connectivity and settlement proximity. In the case of settlement selection the most important variables were urban area, settlement sacral function and local settlement density. The results obtained for roads selection in Poland and settlement selection in Switzerland are promising and demonstrate that the automatic selection approach, developed for Polish data can be applied to the data from other countries. We observed variables that appear to be important for feature selection in both Poland and Switzerland. However, there are also important variables that are specific for each country, providing compelling ground for future research.

Acknowledgement

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."