Optimising road selection for small-scale maps using decision tree-based models

I. Karsznia[a]*, K. Sielicka[a], R. Weibel[b]

[a] Department of Geoinformatics, Cartography and Remote Sensing, Faculty of Geography and Regional Studies, University of Warsaw, Poland
i.karsznia@uw.edu.pl*, k.sielicka@uw.edu.pl
[b] Department of Geography, University of Zurich, Switzerland,
robert.weibel@geo.uzh.ch

Keywords: roads, cartographic generalization, decision trees, small-scale maps, machine learning

Introduction

Reducing the complexity of a road network is necessary after scale change so that the legibility of the map can be maintained. However, deciding about whether to show a particular road section on the map is a very complex process. This process, called selection, constitutes the first operation in the cartographic generalization process. As the first step in the sequence of generalization operations including road simplification, smoothing and displacement, it is a prerequisite to effective road network generalization (Liu et al. 2010).

Generalization of the road network has found the attention of scientists for decades (De Serres and Roy 1990; Yu 2001; Zhang and Li 2011; Weiss and Weibel 2014). Various methods for selecting roads on large scale maps have been developed (e.g. Benz and Weibel 2014), but there is still a lack of effective solutions for small scales. Thus, in this research we concentrate on small scales, however, we also believe that after appropriate modifications, the proposed approach can be applied at large scales.

Researchers agree that an adequate road network generalization process requires considering numerous semantic, geometric, topological and statistical road network characteristics (Richardson and Thomson 1996; Jiang and Claramunt, 2004; Liu et al., 2010; Touya, 2010). For the road selection process, various approaches have been proposed, including graph-theory-based methods, stroke-based methods, methods based on information theory as well as various other measures. However, graph-theory-based methods usually do not consider semantic and geometric roads characteristics, while stroke-based methods rarely take into account statistical road characteristics (Liu et al., 2010; Yan, 2019). Experienced cartographers make decisions based on many conditions and dependencies while simultaneously considering various object characteristics. An effective and automatic algorithm holistically elaborating essential road characteristics would help to reduce the costs of map design, while making it more efficient and faster. Developing such advanced algorithms requires using tools that allow processing large data sets and searching for relevant regularities. Machine learning (ML), successfully used in cartography as well as in many other domains, provides such opportunities. This approach has proven to be a promising solution for settlement selection at small scales (Karsznia and Weibel, 2018; Karsznia and Sielicka, 2019; Karsznia and Sielicka, 2020), generalization of buildings (Sester et al. 2018; Feng et al. 2019), and smoothing and selecting line objects (Lagrange et al. 2000; Balboa and López, 2008; Zhou and Li, 2014).

The paper presents our initial experience of using ML, specifically decision tree-based models, to optimize the selection of the roads from 1:250 000 to 1:500 000 scale. The scope of the research covers designing and verifying road selection models in three selected districts of Poland. The idea was to consider the problem of road generalization holistically, including numerous semantic, geometric, topological and statistical road characteristics.

Method

The source data constitutes the road network contained in the General Geographic Object Database (GGOD) corresponding to 1: 250 000 scale. In the study, three test areas were considered, namely three Polish districts that are different in terms of road network characteristics. Two approaches have been designed. In the basic approach, selection rules were obtained from the regulation of the Polish Ministry of Interior and Administration (Regulation 2011) and applied on the GGOD road data. In the enhanced approach the following steps have been conducted: 1) collecting cartographic knowledge, 2) enriching roads with additional characteristics, 3) formalizing variable values, 4) using machine learning with decision trees, 5) implementing developed models, 6) validating the results.

Within the first step, consultations with experienced cartographers were conducted. Issues such as road characteristics, relations with other map objects, road network continuity and road network patterns maintenance were discussed. In step two, the list of essential road characteristics was gathered. In this study, the characteristics measurably formalized are called variables. The variables concerned both road attributes and spatial characteristics. In step three, the roads were enriched with variables, which give information about their technical parameters.

We also investigated the relations with settlements and other sections of the network, for instance - road network density within the district, number of connections with other roads, centrality measures, type of surface and number of lanes. These variables had to be measurable and comparable among roads constituting the road network. Enriching the road data with relevant variables and considering sufficiently large samples allowed for designing the machine learning models.

For step four above, ML was executed in RapidMiner Studio. The selection process consisted of developing and applying the decision tree-based models. The roads shown on the atlas maps designed by experienced cartographers were used as the training material. As a result, we obtained a decision tree informing which variables are decisive when selecting particular roads. Based on the model derived from the decision tree, the prediction process was conducted. Then, the roads meeting the model requirements were selected. In both approaches, the correctness of selection was assessed by comparing the decisions coming from the automated process to the roads shown on the atlas maps designed by cartographers.

Results

The research results are decision tree-based models and the roads generalized to 1: 500 000 scale. From the obtained decision trees, the most significant variables in the road selection process can be seen. They are placed at the root of the tree (Figure 1).

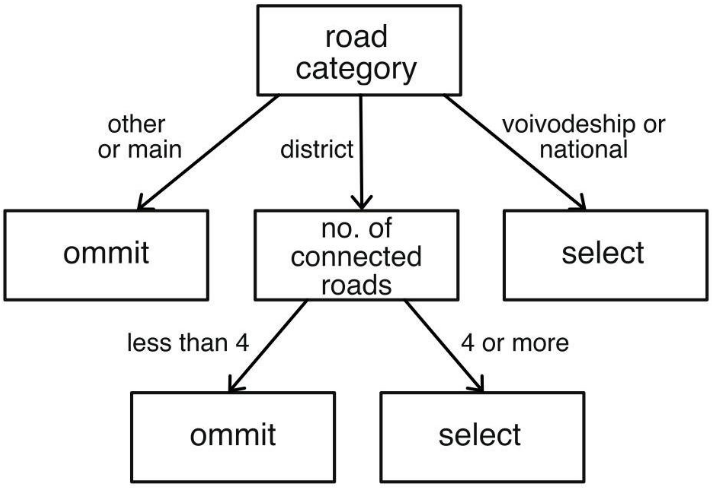

Figure 1: The decision tree for the three districts, result of machine learning.

In the case of the three considered districts the variables that appear on the tree are road management category (road category) and number of road connections (no. of connected roads). Regarding that, by analyzing the tree from the root to its leaves, the road selection rules can be read out. The decision tree obtained for three districts has quite a simple structure (only three decision levels). However, we assume that once we expand our analysis considering more districts, this may make the tree more complex. The accuracy of the obtained results was determined by assessing the similarity percentage between the achieved results and the atlas map designed by the cartographer and was calculated as the number of road segments classified as selected or omitted both on atlas map and in our results. The results of the enhanced approach were also visually compared both to the atlas map and to the results of the basic approach in order to assess network consistency and maintenance of patterns. Figure 2 provides an example of the results for the Bialostocki district.

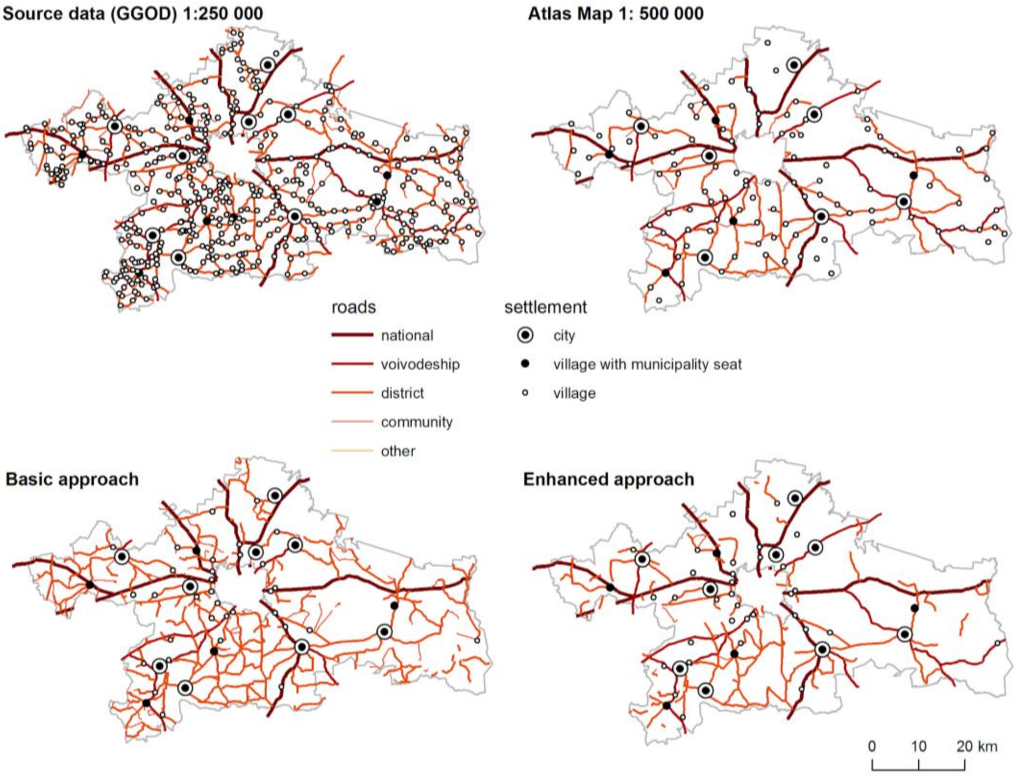

Figure 2: Selection results in Bialostocki district.

Discussion and Conclusion

The visual assessment of the results obtained from the enhanced approach, basic approach and read out from the atlas map is in favor of the enhanced approach (see example of Figure 2). In this approach, the main roads have been preserved, the selected roads constitute a more coherent network, and the general character and differences in the road density are well-reflected. Meanwhile, in the basic approach the road network is too dense compared to the atlas map, and also discontinuous in many places.

The models developed as a result of ML made it possible to improve the accuracy of selection compared to the solution applied in the basic approach. In the case of all tested districts, the accuracy of the enhanced approach is higher than in the case of the basic approach, either while considering all districts or considering each district separately (Table 1). The differences in accuracy are quite significant.

area	basic approach	enhanced approach	difference
all districts	45,10 %	64,61 %	19,51 %
Białostocki	43,70 %	61,19 %	17,49 %

Rzeszowski	55,25 %	72,97 %	17,72 %
Kępiński	42,10 %	65,35 %	23,25 %

Table 1: Selection accuracy for 1:500 000 scale

Finally, one should note that the ultimate goal here is to examine ways to help reduce the cost of map design while making it more efficient and faster vs. complete reconstruction of a manual cartographer's work, because the manual map design process is subjective and may differ depending on the map designer. The fact that the accuracy did not reach 100% means that further work on optimizing the road selection models is recommended. Thus, in future work, using more variables and spreading our studies to more extensive test areas should be considered. By expanding this research, we expect to obtain more complex, but at the same time more informative and holistic decision trees. Valuable steps in future research would include thorough evaluation of the achieved results with the support of experienced cartographers.

References

Balboa, J. L. G., & López, F. J. A., (2008). Generalization-oriented road line classification by means of an artificial neural network. Geoinformatica 12, 289–312. doi: 10.1007/s10707-007-0026-z

Benz, S.A., & Weibel, R. (2014). Road network selection for medium scales using an extended stroke-mesh combination algorithm. Cartography and Geographic Information Science, 41:4, 323-339. doi: 10.1080/15230406.2014.928482

De Serres, B., & Roy, A.G., (1990). Flow direction and branching geometry at junctions in Dendritic River Networks, The Professional Geographer, 42(2) 149–201. doi: 10.1111/j.0033-0124.1990.00194.x

Feng, Y., Thiemann, F., & Sester, M., (2019). Learning Cartographic Building Generalization with Deep Convolutional Neural Network. International Journal of Geo-Information, 8(6), 258. doi: 10.3390/ijgi8060258

Jiang B., & Claramunt C., (2004). A Structural Approach to the Model Generalization of an Urban Street Network. GeoInformatica, 8(2), 157–171. doi: 10.1023/b:gein.0000017746.44824.70

Karsznia, I., & Weibel R., (2018). Improving Settlement Selection for Small-scale Maps Using Data Enrichment and Machine Learning, Cartography and Geographic Information Science, 1-17. doi: 10.1080/15230406.2016.1274237

Karsznia, I., & Sielicka, K., (2019). Exploring essential variables in the settlement selection for small-scale maps using machine learning. Abstracts of the International Cartographic Association, Editor: H. Fujita, vol. 1, 162. doi: 10.5194/ica-abs-1-162-2019.

Karsznia, I., & Sielicka, K., (2020). When Traditional Selection Fails: How to Improve Settlement Selection for Small-Scale Maps Using Machine Learning. ISPRS International Journal of Geo-Information 9(4), 230; doi: 10.3390/ijgi9040230

Lagrange, F., Landras, B., & Mustiere, S. (2000). Machine Learning Techniques for Determining Parameters of Cartographic Generalisation Algorithms. XIXth ISPRS Congress, Vol. XXXIII, Part B4, 718-25. Amsterdam, Netherlands.

Liu, X., Zhan, B., & Ai, T., (2010). Road selection based on Voronoi diagrams and “strokes” in map generalization. International Journal of Applied Earth Observation and Geoinformation, 12, Supplement 2, 194-202. doi: 10.1016/j.jag.2009.10.009

Regulation of the Ministry of Interior on 17 November 2011 on the Topographic Objects Database and General Geographic Objects Database, as well as standard cartographic products, Journal of Laws of 2011, No 279 item 1642.

Richardson, D. E., & Thomson, R.C., (1996). Integrating Thematic, Geometric, and Topological Information in the Generalization of Road Networks. Cartographica: The International Journal for Geographic Information and Geovisualization, 33(1), 75-83. doi: 10.3138/F150-7678-5Q15-8N06

Sester, M., Feng, Y., & Thiemann, F., (2018). Building generalization using deep learning. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-4, 2018 ISPRS TC IV Mid-term Symposium “3D Spatial Information Science – The Engine of Change”, 1–5 October 2018, Delft, The Netherlands.

Touya, G., (2010). A road network selection process based on data enrichment and structure detection. Transactions in GIS, 14 (5), 595–614. doi: 10.1111/j.1467-9671.2010.01215.x

Weiss, R., & Weibel, R., (2014). Road network selection for small-scale maps using an improved centrality-based algorithm. Journal of Spatial Information Science, 9, 71-99. doi: 10.5311/JOSIS.2014.9.166

Yan, H., (2019). Description Approaches and Automated Generalization Algorithms for Groups of Map Objects. Springer Publishing Company

Yu, X., (2001). Road Network simplification with knowledge-based spatial analysis. Supplement Journal of Geographical Sciences, 11, 54–62.

Zhang, H., & Li, Z., (2011). Weighted ego network for forming hierarchical structure of road networks. International Journal of Geographical Information Science, 25(2), 255–272. doi: 10.1080/13658810903313534

Zhou Q., Li Z., (2014). Use of Artificial Neural Networks for Selective Omission in Updating Road Networks. The Cartographic Journal 51(1), 38-51. doi: 10.1179/1743277413Y.0000000042