

Summary

In the past decade, the process of cartographic generalization has undergone significant changes. This is due to the increasing role of artificial intelligence (AI), including classical machine learning (ML) and deep learning (DL), in the generalization process. Recent studies even refer to deep learning as a new paradigm in cartographic generalization, fundamentally improving the process by enabling its full automation and optimization. To follow this paradigm requires new principles of generalization, along with a better understanding of how users perceive and interpret AI-generated maps, and a thorough evaluation of such maps. These efforts are essential to verify existing and develop new rules for map design and generalization.

This monograph focuses on a key stage of the cartographic generalization process that is selection. It is the first and most fundamental, though not the only, stage of generalization. After selection, further stages follow, such as quantitative and qualitative generalization, including aggregation (merging), simplification, and smoothing. The research presented in this paper concerns three selected, yet most important, thematic layers that constitute the content of a topographic map: settlements represented by point symbols, the road network, and the river network. Naturally, a complete map also includes settlement outlines, forest outlines, lakes, relief, and other features. These, however, are not part of the present study but should be considered in future research to propose more holistic solutions encompassing the full map content.

The direct inspiration and motivation for undertaking this research project stem from a lack of comprehensive solutions based on AI and of a properly formalized cartographic knowledge base that could be implemented and utilized in a computational environment. This project aims to fill that gap through one of the first attempts to optimize the design of basic topographic map content in a manner that is (1) fully automated, (2) based on a formalized cartographic knowledge base, and thus optimal, and (3) utilizing state-of-the-art AI, ML, and DL algorithms.

The study focused on the automatic selection of settlements, roads, and rivers—the main thematic categories of the Database of General Geographic Objects (BDOO)—generalized from the 1:250,000 level of detail to 1:500,000 and 1:1,000,000

scales. The research employed artificial intelligence techniques, particularly selected machine learning models and elements of graph theory.

The study is both conceptual and methodological in nature, as it involves the development of methods, but also the implementation and verification of the proposed approaches. To explore, understand, and automate the selection process, several machine learning models were implemented and tested at different stages of the research: Decision Trees (DT), Decision Trees with Genetic Algorithms (DT-GA), Support Vector Machines (SVM), Random Forest (RF), Artificial Neural Networks (ANN), and Deep Learning (DL).

The research is of an interdisciplinary character, combining elements of cartography, geoinformatics, computer science (including data science), and socio-economic geography. The analyzed features include both anthropogenic elements (settlements, the transportation network) and natural ones (the river network as a hydrological element). Given its interdisciplinary nature and thematic scope, the field of Socio-Economic Geography and Spatial Management provides an optimal and valuable conceptual framework for these analyses.

The main outcome of this research is a methodology for the automatic selection of settlements, roads, and rivers, developed to optimize map design in small-scale cartography using selected machine learning models. The proposed methodology enables full automation of the selection process from the 1:250,000 scale to 1:500,000 and 1:1,000,000. Within this framework, key attributes of settlements, roads, and rivers were identified as critical decision variables in the selection process. The use of machine learning models also made it possible to explore and define additional selection rules, which serve as an important—and often necessary—extension of the principles contained in current map editing guidelines. Furthermore, the methodology comprises an analysis of the importance, weights, and correlations of the variables used in the selection process.

This monograph consists of introduction, five chapters and conclusion. The introduction defines the objectives and scope of the research and outlines the main findings. The study focuses on cartographic generalization and artificial intelligence, whose definitions and classifications—consistent with the current state of knowledge—are presented in Chapter 1. This chapter also discusses the state of research on the automation of cartographic generalization and introduces the concept of artificial intelligence and its subfield, machine learning, along with the main types of ML methods. It also defines cartographic generalization as related to AI, explaining the role of one of its key operations—selection, i.e., the inclusion or omission of particular map features—and its function in modern generalization models. Additionally, sources of cartographic knowledge necessary for data generalization are identified, including their acquisition, formalization, and implementation. Chapter 2 reviews previous attempts to apply machine learning to cartographic generalization, categorizing them by feature type: settlements, road networks, and river networks. Chapter 3 characterizes the machine learning models used in this research. Chapter 4 describes the subject and scope of the study, while Chapter 5 presents a detailed

methodology for the automatic selection of settlements, roads, and rivers, including its implementation through programming scripts, developed ML models, and algorithms. The monograph concludes with a synthesis of the research results, emphasizing its potential applications and implementation possibilities, as well as outlining directions for future work.

Keywords: cartographic generalization, artificial intelligence (AI), machine learning (ML), automation, knowledge formalization, small scales, settlements, roads, rivers, deep learning (DL)

Streszczenie

W ostatniej dekadzie proces generalizacji kartograficznej przeszedł istotne zmiany, głównie za sprawą rosnącej roli sztucznej inteligencji (SI), w tym klasycznego uczenia maszynowego (UM) oraz uczenia głębokiego (UG). Najnowsze badania określają uczenie głębokie wręcz jako nowy paradygmat w generalizacji kartograficznej, który zasadniczo usprawnia ten proces poprzez umożliwienie jego pełnej automatyzacji i optymalizacji. Aby realizować ten paradygmat, potrzebne są nowe zasady generalizacji, lepsze zrozumienie sposobu, w jaki użytkownicy postrzegają i interpretują mapy generowane przez AI, a także dokładna ich ocena. Wysiłki te są niezbędne, by weryfikować istniejące oraz opracowywać nowe reguły projektowania i generalizacji map.

W monografii skoncentrowano się na kluczowym etapie generalizacji kartograficznej – selekcji. Jest to etap pierwszy i podstawowy, choć nie jedyny. Po selekcji następują kolejne etapy, takie jak generalizacja ilościowa i jakościowa, w tym agregacja (łączenie), uproszczenie i wygładzanie. Przedstawione w pracy badania dotyczą trzech wybranych i najważniejszych warstw tematycznych, które tworzą treść mapy topograficznej: miejscowości (reprezentowanych symbolami punktowymi), sieci drogowej oraz sieci rzecznej. Oczywiście pełna treść mapy obejmuje również granice administracyjne, kontury lasów, jeziora, rzeźbę terenu i inne elementy. Te jednak nie zostały uwzględnione w opracowaniu, lecz powinny zostać objęte przyszłymi badaniami w celu zaproponowania bardziej kompleksowych rozwiązań uwzględniających pełną treść mapy.

Bezpośrednią inspiracją i motywacją do podjęcia badań był brak kompleksowych rozwiązań opartych na AI oraz brak odpowiednio sformalizowanej bazy wiedzy kartograficznej, którą można byłoby zaimplementować i wykorzystać w środowisku obliczeniowym. Celem badań jest wypełnienie tej luki poprzez jedną z pierwszych prób optymalizacji projektowania podstawowej treści mapy topograficznej w sposób (1) w pełni zautomatyzowany, (2) oparty na sformalizowanej bazie wiedzy kartograficznej, a zatem optymalny, oraz (3) wykorzystujący najnowocześniejsze algorytmy SI, UM i UG.

W badaniach skoncentrowano się na automatycznej selekcji miejscowości, dróg i rzek – głównych kategorii tematycznych Bazy Danych Obiektów Ogólnogeograficznych

(BDOO) – z poziomu szczegółowości 1:250 000 do skal 1:500 000 oraz 1:1 000 000. W badaniach zastosowano techniki sztucznej inteligencji, w szczególności wybrane modele uczenia maszynowego oraz elementy teorii grafów.

Praca ma charakter zarówno koncepcyjny, jak i metodologiczny – opracowano metody, ale również przetestowano ich implementację i skuteczność. W celu zbadania, zrozumienia i automatyzacji procesu selekcji zaimplementowano i przetestowano kilka modeli ML na różnych etapach badań: drzewa decyzyjne (DT), drzewa decyzyjne z algorytmami genetycznymi (DT-GA), maszynę wektorów nośnych (SVM), las losowy (RF), sztuczne sieci neuronowe (ANN) oraz modele uczenia głębokiego (UG).

Badania mają charakter interdyscyplinarny, łączą elementy kartografii, geoinformatyki, informatyki (w tym *data science*) oraz geografii społeczno-ekonomicznej. Analizowane obiekty obejmują elementy zarówno antropogeniczne (miejscowości, sieć transportową), jak i przyrodnicze (sieć rzeczna jako element hydrologiczny). Ze względu na interdyscyplinarność oraz zakres tematyczny optymalną i wartościową ramę koncepcyjną dla zawartych w pracy analiz stanowi dziedzina geografii społeczno-ekonomicznej i gospodarki przestrzennej.

Głównym rezultatem badań jest metodologia automatycznej selekcji miejscowości, dróg i rzek opracowana w celu optymalizacji redakcji map w małych skalach z wykorzystaniem wybranych modeli uczenia maszynowego. Proponowana metodologia umożliwi pełną automatyzację procesu selekcji w przejściu od skali 1:250 000 do skal 1:500 000 i 1:1 000 000. W jej ramach zidentyfikowano kluczowe atrybuty miejscowości, dróg i rzek jako istotne zmienne decyzyjne. Wykorzystanie modeli ML pozwoliło również na eksplorację i sformułowanie dodatkowych reguł selekcji, które stanowią ważne – a często niezbędne – uzupełnienie zasad zawartych w obowiązujących wytycznych redakcyjnych map. Metodologia zawiera także analizę istotności, wag i korelacji zmiennych użytych w procesie selekcji.

Monografia składa się z wprowadzenia, pięciu rozdziałów i podsumowania. We wprowadzeniu określono cele i zakres badań oraz przedstawiono główne wnioski. W pracy skupiono się na generalizacji kartograficznej i sztucznej inteligencji, których definicje i klasyfikacje – zgodne ze stanem aktualnej wiedzy – zostały przedstawione w rozdziale pierwszym. Omówiono w nim również stan badań nad automatyzacją generalizacji kartograficznej oraz wprowadzono pojęcie uczenia maszynowego jako poddziedziny SI – wraz z głównymi typami metod UM. Zdefiniowano także generalizację kartograficzną w odniesieniu do SI, wyjaśniając rolę jednej z jej kluczowych operacji – selekcji, czyli wyboru lub pominięcia elementów mapy – oraz jej funkcję w nowoczesnych modelach generalizacji. Zidentyfikowano źródła wiedzy kartograficznej niezbędne do generalizacji danych, w tym ich pozyskiwanie, formalizację i implementację. W rozdziale drugim dokonano przeglądu dotychczasowych prób zastosowania UM do generalizacji kartograficznej, sklasyfikowanych według typu obiektu: miejscowości, sieć drogowa i sieć rzeczna. W rozdziale trzecim scharakteryzowano modele UM wykorzystane w badaniach. W rozdziale czwartym opisano przedmiot i zakres badań, a w rozdziale piątym zaprezentowano szczegółową metodologię automatycznej selekcji miejscowości, dróg i rzek, wraz z jej implementacją

w postaci skryptów programistycznych, opracowanych modeli UM i algorytmów. Monografia kończy się podsumowaniem, w którym zawarto syntezę badań, wskazując potencjalne zastosowania i możliwości wdrożenia, a także zarysowując kierunki dalszych prac.

Słowa kluczowe: generalizacja kartograficzna, sztuczna inteligencja (SI), uczenie maszynowe (UM), automatyzacja, formalizacja wiedzy kartograficznej, skale przeglądowe, miejscowości, drogi, rzeki, uczenie głębokie (UG)