Advances in Science, Technology & Innovation

IEREK Interdisciplinary Series for Sustainable Development

Advances in Science, Technology & Innovation (ASTI) is a series of peer-reviewed books based on the best studies on emerging research that redefines existing disciplinary boundaries in science, technology and innovation (STI) in order to develop integrated concepts for sustainable development. The series is mainly based on the best research papers from various IEREK and other international conferences, and is intended to promote the creation and development of viable solutions for a sustainable future and a positive societal transformation with the help of integrated and innovative science-based approaches. Offering interdisciplinary coverage, the series presents innovative approaches and highlights how they can best support both the economic and sustainable development for the welfare of all societies. In particular, the series includes conceptual and empirical contributions from different interrelated fields of science, technology and innovation that focus on providing practical solutions to ensure food, water and energy security. It also presents new case studies offering concrete examples of how to resolve sustainable urbanization and environmental issues. The series is addressed to professionals in research and teaching, consultancies and industry, and government and international organizations. Published in collaboration with IEREK, the ASTI series will acquaint readers with essential new studies in STI for sustainable development.

More information about this series at http://www.springer.com/series/15883

Biswajeet Pradhan · Maher Ibrahim Sameen

Laser Scanning Systems in Highway and Safety Assessment

Analysis of Highway Geometry and Safety Using LiDAR

Springer

Biswajeet Pradhan
Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS)
University of Technology Sydney
Sydney, Australia

Maher Ibrahim Sameen
Centre for Advanced Modelling and Geospatial Information Systems (CAMGIS)
University of Technology Sydney
Sydney, Australia

ISSN 2522-8714 ISSN 2522-8722 (electronic)
Advances in Science, Technology & Innovation
IEREK Interdisciplinary Series for Sustainable Development
ISBN 978-3-030-10373-6 ISBN 978-3-030-10374-3 (eBook)
https://doi.org/10.1007/978-3-030-10374-3

Library of Congress Control Number: 2019933705

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Recent developments in laser scanning technologies have provided innovative solutions for acquiring three-dimensional (3D) point clouds about road corridors and its environments. Unlike traditional field surveying, satellite imagery, and aerial photography, laser scanning systems offer unique solutions for collecting dense point clouds with millimeter accuracy and in a reasonable time. The data acquired by laser scanning systems empower modeling road geometry and delineating road design parameters such as slope, superelevation, and vertical and horizontal alignments. These geometric parameters have several geospatial applications such as road safety management.

The purpose of this book is to promote the core understanding of suitable geospatial tools and techniques for modeling of road traffic accidents by the state-of-the-art artificial intelligence (AI) approaches such as neural networks (NNs) and deep learning (DL) using traffic information and road geometry delineated from laser scanning data.

Data collection and management in databases play a major role in modeling and developing predictive tools. Therefore, the first two chapters of this book introduce laser scanning technology with creative explanation and graphical illustrations and review the recent methods of extracting geometric road parameters. The third and fourth chapters present an optimization of support vector machine and ensemble tree methods as well as novel hierarchical object-based methods for extracting road geometry from laser scanning point clouds.

Information about historical traffic accidents and their circumstances, traffic (volume, type of vehicles), road features (grade, superelevation, curve radius, lane width, speed limit, etc.) pertains to what is observed to exist on road segments or road intersections. Soft computing models such as neural networks are advanced modeling methods that can be related to traffic and road features to the historical accidents and generates regression equations that can be used in various phases of road safety management cycle. The regression equations produced by NN can identify unsafe road segments, estimate how much safety has changed following a change in design, and quantify the effects of road geometric features and traffic information on road safety.

This book aims to help graduate students, professionals, decision makers, and road planners in developing better traffic accident prediction models using advanced neural networks.

This book is organized into twelve chapters.

Chapter 1 presents an overview of LiDAR technology for geometric road modeling. The concepts of airborne, mobile, and terrestrial systems were explained. The strengths and weaknesses of these systems were discussed in general and in the context of road modeling. The detail literature review indicates that the three LiDAR systems (ALS, MLS, and TLS) share similarity in hardware components (GNSS, IMU, and laser scanner), range measurements principles, and supporting digital cameras. In contrast, the literature showed that these systems differ in accuracy, the safety of data collection, and overall efficiency. Regarding road geometry modeling, the ground-based systems are easier to operate than ALS because the latter requires integration of IMU and GNSS sensors for accurate point georeferencing. Ground-based systems offer more options for setup locations including away from the road. Additionally, MLS and TLS systems provide significantly improved horizontal accuracy due to looking angle. However, ALS has a better view of moderately sloping roadside features or flat terrain such as of the pavement surface compared with ground-based systems depending on their scanner orientation. Features such as ditches and road barriers can be better viewed by ALS, whereas other systems are more appropriate to model the cliff slopes. Over the last decades, the development of LiDAR technology focused on increased measurements rates improved positional accuracy and angular precision and improvements in instrument functionality and easy to operate. The continuous developments of LiDAR will likely result in better road geometry modeling. This is because, with more accurate and denser point clouds, more detailed road features can be extracted and modeled. However, information such as road material types and surface roughness may require additional ancillary data such as multispectral and hyperspectral images. Overall, the ground-based systems are preferable to airborne systems for road geometry modeling due to the reduction in occlusion. In densely vegetated areas, the airborne LiDAR system is more likely to fail to collect detailed road features such as road markings. However, road geometry modeling using MLS and TLS data requires more sophisticated and efficient algorithms than those for ALS data.

Chapter 2 presents a review of extraction and geometric modeling of road networks using LiDAR data. A brief explanation with graphical illustrations was given for geometric road models including their design and asset elements. This chapter also provided a novel classification of geometric road modeling according to the sensor type, preset objectives, and the technique used for data processing. Lastly, it discussed the current challenges and future trends in extraction and geometric modeling of road networks using LiDAR point clouds. The current algorithms do not meet the increasing demands and requirements of industrial applications. One key solution to this challenge was the automation of extraction and geometric modeling of road networks. However, most of the modern roads are complex in geometry, which by using automatic data processing methods; the detection accuracy is often not satisfactory. In addition, the presence of occlusions due to cars, trees, and buildings generates extra challenges and degrades the accuracy of the models. As a result, the future directions of extraction and geometric modeling of road networks should focus on developing models that can best solve the above issues.

Chapter 3 discusses the recent technologies in mobile mapping systems that have enabled the rapid and cost-effective data acquisition on road corridors. This chapter presents a novel hierarchical strategy for the semiautomatic extraction of geometric road parameters from mobile laser scanning data. The hierarchical strategy included mean shift segmentation, particle swarm optimization, support vector machine classification, and principal component analysis. Accuracy assessments showed that the lowest and highest errors of the slope parameter were 1.15 and 28.57%, respectively, whereas for the superelevation parameter was 0.96 and −13.50 %, respectively. These values of errors and comparative studies demonstrate the effectiveness of the method and indicate that the proposed model can offer satisfactory results.

Chapter 4 presents optimization of parameters in extracting features from laser scanning data. Optimization is one of the essential tasks in machine learning algorithms. Classification algorithms, such as SVM and ensemble tree, require several user-defined parameters for model calibration and data classification. In this study, Taguchi-based optimization was implemented to select the best combination of user-defined parameters for the SVM and ensemble tree algorithms to extract road networks from LiDAR and aerial orthophoto data. The accuracy matrices indicated that the AUC values for the SVM and ensemble tree methods were 0.71 and 0.89, respectively, when the default parameters were used. Meanwhile, when the optimized parameters were used in classification, the AUC values of the SVM and ensemble tree methods were 0.88 and 0.95, respectively. This research demonstrates that optimizing the user-defined parameters of classification methods is necessary to improve accuracy and quality in detecting and classifying road networks from LiDAR data. In this study, the selected algorithms were trained using randomly selected samples and then tested on raster images. Further evaluation of the optimization-based classification is required to measure the efficiency and reliability of the approach. In addition, testing the transferability of the developed models on different study areas and datasets is a dynamic research trend.

Chapter 5 presents an accurate approach for extraction of highway information from remote sensing data. This is significant for various applications such as traffic accident modeling, navigation, intelligent transportation systems, and natural hazard assessments. One of the conventional technique used for automatic highway extraction is machine learning. Despite several machine learning algorithms have been tested and tried in the recent years; however, there is no common agreement has been made which method performs better and spatially transferable. Therefore, this paper contributes in evaluating several machine learning algorithms (i.e. support vector machine, logistic regression, neural network, and decision tree) for automatic highway extraction from high-resolution airborne LiDAR data. Based on the comparative study performed, the best among studied machine learning algorithms was identified and used in an integrated GIS workflow for automatic highway extraction. The advanced GIS integrated workflow is an efficient model that could be applied to most commercial and open-source GIS software. Among the studied machine learning algorithms, multilayer perceptron and decision tree algorithms showed the best overall accuracy tested on randomly selected sampling data. However, when the transferability of the models investigated, logistic regression found to be the optimal algorithm for highway extraction from LiDAR data. In addition, although support vector machine produced high overall accuracy (90.19%) on sampling data, the model produced low-quality classification when applied to raster data. Thus, it suffers from model transferability issues. The quantitative evaluation showed that the logistic regression model could extract highway features from LiDAR data with 85.43% for completeness measure, 76.70%, and 67.82% for correctness and quality measures, respectively. The result of this study provides a clear guideline for other researchers to develop more advanced and automatic GIS models for accurate extraction of highways from LiDAR data.

Chapter 6 discusses the several factors that contribute to road traffic accidents including human, vehicle, road geometry, and environmental factors. This chapter discusses the impact of roadside features (e.g., trees, access points, median, and shoulder) that were extracted from a mobile laser scanning (MLS) data on the injury severity of road traffic accidents. The analysis and the discussions were based on a case study of North–South Expressway (NSE) in Malaysia. The proposed methodology for roadside feature delineation was based on an object-based analysis comprised of multiresolution segmentation, optimization of features by random forest, and classification with support vector machines. Logistic regression (LR) model was utilized to analyze the correlation between the extracted features and the accident records from 2009 to 2015. The results of this study show that the proposed methodology could extract the roadside features from the MLS data with an overall accuracy of 86.45%. In addition, the estimated errors in calculating the road median, road width, and road shoulder width were 0.15, 0.31, and 0.45 m, respectively. On the other hand, results of LR revealed that shoulder width, high density of trees, poor lighting conditions, and involving motorcycle in crashes increase the injury severity of the accidents. Therefore, to improve the road safety in the focused area, this study suggests that these factors should be considered in road maintenance, safety management, planning, and other transportation projects.

Chapter 7 presents a novel GIS-based model for automatic identification of road geometry in vector data. Some geospatial applications such as road safety assessment, car navigation, and updating digital road maps usually require road geometry information. In this research, a new model based on geographic information system (GIS) was proposed to automate the process of identifying road geometry in vector polylines. The proposed model first applies a Bezier interpolation to smooth the polylines for better cartographic representation. Then, the polylines were converted into raster data at 0.5 m spatial resolution. This data conversion enabled to convert the polylines to a set of points. After that, three geometry predictors were estimated from the set of points, point density, length of a line segment, and a cumulative angle between five consecutive points. Finally, the geometry predictors were used to predict the road geometry using three classification methods, namely support vector machine (SVM), decision tree (DT), and logistic regression (LR). The results show that the proposed model could identify road geometry in vector data. Overall accuracies of the three classification methods were 87.4, 89.7, and 87.2%, respectively; DT is the best.

Chapter 8 reviews methods used to model the frequency and injury severity of traffic accidents. First, a general overview of modeling of traffic accidents was given. Second, road accident setting was described, which included configurations of road accidents across different studies. Some works researched traffic accidents on traffic signals, plaza tolls, whereas others focused on highways and main roads. Third, model factors for predicting the frequency and injury severity of traffic accidents were reviewed and discussed. Briefly, the selection of model factors depends on the purpose of research, the data availability, and the definition of the accident frequency or injury severity. In general, the reviewed studies showed that the model factors used for predicting road accident severity are usually more in number as compared to model factors used for predicting road accident frequency. After that, since most of the works showed that traffic accident data could be modeled by two primary models, statistical and computational intelligence, they were reviewed and compared considering their differences, similarities, advantages, and disadvantages. Finally, NN and DL models were further examined in details reviewing their concepts, use, the purpose of usage, and types. Besides the traditional NN model, RNN and CNN models were also discussed. In general, the accuracy of NN model decreases in particular with the increase in complexity of prediction, as is the case with road accident severity. The predictive accuracy of NN for accidents severity can be increased with the help of fusion algorithms and a clustering method. In addition, using a larger number of input nodes in the NN structure and more hidden layer, the predictive accuracy of NN is reported to be increased.

Chapter 9 presents a comparative assessment of neural networks and support vector machines in modeling traffic accident severity using actual reported causes. With the significant increase in urbanization and number of registered vehicles worldwide, traffic accidents are a global concern. This chapter presents an evaluation of deep neural networks (DNNs) and support vector machines (SVMs) for traffic accident severity modeling using reported causes. A case study of Malaysian North–South Expressway (NSE) was presented. Accident data for the period of 2009–2015 consisting of 1138 observations were acquired for the study area. The data contained seven explanatory variables (vehicle type, accident cause, collision type, lighting condition, zone, bound, and road surface condition). The data were split into three subsets as follow: (1) training set, which contained 70% of the entire data; (2) validation set contained 15% of the data; and (3) testing set contained the remaining 15%. Using the training and validation datasets, the DNN and SVM models with optimum hyperparameters were built. After that, the trained models were tested using the testing dataset. The results indicated that the linear SVM outperformed other SVM models and NN models. The highest accuracy on testing dataset was 71.34% obtained by the linear SVM model. In addition, the best DNN model obtained 70.67%. Furthermore, the random forest analysis showed that the reported accident cause and vehicle type are the two most influential factors increasing the severity of accidents in the study area.

Modeling the injury severity of road traffic accidents is a challenging yet essential task in transportation management. Predictive models should be powerful enough to make accurate forecasts and generalize well on different data subsets for a given area. Chapter 10 presents a new hybrid algorithm based on extreme gradient boosting (XGBoost) and deep neural networks (DNNs) for predicting the injury severity of road traffic accidents. The proposed model workflow includes handling imbalance data, optimizing the input factors, fine-tuning the model's hyperparameters by a grid search method, and combining two dominant models XGBoost and DNN in a single hybrid model to improve the generalization performance. The findings show that the handling imbalance data with the SMOTEENN method can improve the model performance significantly ($\sim$by 12%). In addition, feature selection and optimization of the hyperparameters could help improving the model regarding the prediction accuracy and generalization. The new hybrid model had an average accuracy of 0.96, which outperformed many other machine learning models like K-nearest neighbors, decision trees, random forest, multilayer perceptron, support vector machine, naïve Bayes, Gaussian process classifier, and quadratic discriminant analysis methods. On the other hand, the XGBoost model with optimum features also performed well (0.95) slightly worse than the hybrid model. The new models allow better road safety assessment, and with their high predictive and generalization capacity, they provide a better understanding of road accidents on highways. Information created from such advanced models can help decision makers to take better decisions eliminating the significant losses due to accidents in future.

Traffic accidents are becoming a growing concern due to the increasing number of casualties and losses in economic activities. Driver, highway, vehicle, accident characteristics, and climatic factors are some of the factors that determine the severity level. These factors are often analyzed using statistical and machine learning models. In recent years, new advances in computer hardware/software engineering and large datasets have assisted deep learning to win numerous contests in pattern recognition, machine learning, and remote sensing. Deep learning enables hierarchal learning by computational models of data using multiple abstraction levels at different processing layers. Chapter 11 investigates the power of deep learning in predicting the severity of injuries when accidents occur due to traffic on Malaysian highways. Three network architectures based on a simple feedforward neural networks (NNs), recurrent neural networks (RNNs), and convolutional neural networks (CNNs) were proposed and optimized through a grid search optimization to fine-tune the hyperparameters of the models that can best predict the outputs with less computational costs. The results showed that among the tested algorithms, the RNN model with an average accuracy of 73.76% outperformed the NN model (68.79%) and the CNN (70.30%) model based on a tenfold cross-validation approach. On the other hand, the sensitivity analysis indicated that the best optimization algorithm is 'Nadam' in all the three network architectures. In addition, the best batch size for the NN and RNN was determined to be 4 and 8 for the CNN. The dropout with keep probability of 0.2 and 0.5 was found critical for the CNN and RNN models, respectively. This research has shown that deep learning models such as CNN and RNN provide additional information inherent in the raw data such as temporal and spatial correlations that outperform the traditional NN models regarding both accuracy and stability.

Chapter 12 presents a novel deep learning model based on recurrent neural network (RNN) and transfer learning (TL) for predicting injury severity of motorcycle accidents on Malaysians expressways. Due to the limited available data (155 records) on motorcycle accidents, we first developed a model based on 970 accident records that accurately predicts vehicle accidents on Malaysians expressways. Then, this model was retrained on a small dataset of the motorcycle accidents by freezing the first few layers. In addition, the network architecture and its hyperparameters were optimized via grid search method. The best network comprised of one long short-term memory (LSTM) layer (128 hidden units) and two dense layers of 100 hidden units. Moreover, it also contained dropout layers with a probability of 0.2 to reduce the model complexity and improve the generalization performance of the model. The network was trained using RMSprop optimization method with learning rate and weight decay of 0.001 and 0.0001, respectively. The model was implemented using Google's TensorFlow. The proposed model achieved a training and testing accuracy of 78.29 and 77.90%, respectively, and the TL could improve the accuracy by almost 10%. A comparative study showed that the RNN with TL method outperformed other machine learning models such as support vector machine, logistic regression, and random forest. The model also could explain the effects of accident-related factors on the injury severity outcomes by calculating factor's coefficients by a connection weight method. The results suggested that collision type, accident cause, and lighting condition are the most influential factors for increasing the injury severity of motorcycle accidents on Malaysians expressway. Moreover, the proposed model can efficiently use the temporal and spatial structure of the accident data as additional information to improve the prediction performance, which other methods lack.

The publication of this book would not have been possible without an excellent cooperation from my colleagues at Springer Verlag, Germany. Special thanks to Dr. Nabil Khelifi for motivating and encouraging the first author to write this book. At Springer Verlag, the efforts from Reyhaneh Majidi is appreciable.

Sydney, Australia
April 2019

Biswajeet Pradhan

Contents