

Impact Factor:

ISRA (India) = 1.344	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 0.829	PIHII (Russia) = 0.207	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 4.102	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 2.031	

SOI: [1.1/TAS](#) DOI: [10.15863/TAS](#)

International Scientific Journal Theoretical & Applied Science

p-ISSN: 2308-4944 (print) e-ISSN: 2409-0085 (online)

Year: 2018 Issue: 06 Volume: 62

Published: 26.06.2018 <http://T-Science.org>

Irina Ivanovna Vasileva

Student,
Peter the Great St.Petersburg Polytechnic University

Vadim Andreevich Kozhevnikov

Senior Lecturer,
Peter the Great St.Petersburg Polytechnic University

Oleg Yurievich Sabinin

Candidate of Engineering Sciences,
Peter the Great St.Petersburg Polytechnic University

SECTION 4. Computer science, computer engineering and automation.

UDC 004.4

GEOSPATIAL ANALYSIS AND ANDROID DEVELOPMENT FOR CREATING APPLICATIONS WORKING WITH CRIME DATA

Abstract: Article describes the creation of an application for building safe routes using geospatial data analysis methods. Also this article describes the results of predictive algorithms efficiency analysis for the crime rate prediction.

Key words: geospatial analysis, predictive analytics, heatmap, crime analysis, prediction efficiency, software development, Android, Java, Python.

Language: English

Citation: Vasileva II, Kozhevnikov VA, Sabinin OY (2018) GEOSPATIAL ANALYSIS AND ANDROID DEVELOPMENT FOR CREATING APPLICATIONS WORKING WITH CRIME DATA. ISJ Theoretical & Applied Science, 06 (62): 121-126.

Soi: <http://s-o-i.org/1.1/TAS-06-62-22> **Doi:**  <https://dx.doi.org/10.15863/TAS.2018.06.62.22>

Introduction

Nowadays, more and more people use Android smartphones or tablets to find a path from one point to another. Paper maps gave up their place to the convenience which mobile devices can provide.

Today many countries pay attention to digital cartography, which has become a serious tool for improving the quality of life of citizens. The scope of digital maps is quite wide. This area, called geospatial analysis, includes various types of analyses and predictions, such as traffic analysis, population analysis, geomarketing, social mapping, crime analysis etc.

Geospatial analysis is a big topic that combines parts of data analysis and digital mapping. Geospatial analysis can be very useful for planning the infrastructure of the city.

The combination of the convenience of mobile devices and the capabilities that geospatial analysis provides can significantly improve the lives of citizens.

Motivation

The purpose of this work is to develop a system for building safe routes using geospatial data analysis methods.

As it was mentioned above, there are many people who use their mobile devices to find routes.

After analyzing the situation, it turned out that at the moment there are no applications that build routes taking into account the criminal situation in the city of St. Petersburg. Thus, it was decided to create an application that can use crime data and build routes for St. Petersburg.

The main purpose of this application is to reduce the level of crime in the city by creating an opportunity for each user to build safe routes.

Implementation

The implementation is based on two main parts: data analysis and Android application.

Data analysis includes analysis of crime data for 2017, extraction of features for prediction, predictive analysis of crime rate for 2018 and transformation data into data sets of weighted geo points.

To analyze the crime rate in 2017, criminal news from the website of the Ministry of internal Affairs were used. [1]

To extract features for prediction, we used data on amenities and stores from OpenStreetMap for St. Petersburg and open data on apartments in St. Petersburg. [2; 3]

For predictive analysis the following algorithms were used:

Impact Factor:

ISRA (India) = 1.344	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 0.829	PIHII (Russia) = 0.207	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 4.102	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 2.031	

1. Linear regression is the simplest and most classic linear method for regression. Linear regression finds the parameters that minimize the mean squared error between predictions and the true regression targets on the training set. The mean squared error is the sum of the squared differences between the predictions and the true values. [4, p. 47]

2. Kernelized support vector machines (often just referred to as SVMs) are an extension that allows for more complex models that are not defined simply by hyperplanes in the input space. [4, p. 92]

3. Decision trees are widely used models for classification and regression tasks. Essentially, they learn a hierarchy of if/else questions, leading to a decision. [4, p. 70]

4. Random forest is essentially a collection of decision trees, where each tree is slightly different from the others. The idea behind random forests is that each tree might do a relatively good job of predicting, but will likely overfit on part of the data. If we build many trees, all of which work well and overfit in different ways, we can reduce the amount of overfitting by averaging their results. [4, p. 83]

5. Gradient boosting for regression trees is another ensemble method that combines multiple decision trees to create a more powerful model. The main idea behind gradient boosting is to combine many simple models (in this context known as weak learners), like shallow trees. Each tree can only provide good predictions on part of the data, and so more and more trees are added to iteratively improve performance. [4, p. 88-89]

The result of the complex prediction of these algorithms has been transformed into data sets consisting of weighted geographic points. We also transformed crime data for previous periods into weighted data sets to create an analyzing mode in our Android application.

Data analysis was implemented in Python using the machine learning library "scikit-learn". [5; 6]

Android application is based on Google Maps. It creates a heat map based on the weighted data sets from the previous step and finds paths using the Google API. [7]

Heat maps show the density of point features with a yellow-orange-red color continuum. Google's Geo Developers Blog describes these maps as a representation of "geospatial data on a map by using different colors to represent areas with different concentrations of points – showing overall shape and concentration trends". [8]

The application supports two modes: situation analysis and pathfinding.

In the first mode, the user can select the year on the slider from 2010 to 2019, for which a heat map of criminality will be drawn. (Fig.1) This mode allows the user to analyze the situation with crime in the city in retrospect, in the current moment and in the future, which can be useful for planning the infrastructure of the city.

In the second mode, the user can build a safe route between two points. (Fig.2) In this mode user also can see the crime heat map for 2018 year.

It is also worth mentioning that the heat map is interactive and changes its appearance when user scales the map. (Fig.3)

Android application was implemented in Java with Android development techniques. [9; 10]

The interaction between the user and the application is done through activities.

The Activity class is a crucial component of an Android app, and the way activities are launched and put together is a fundamental part of the platform's application model. [10]

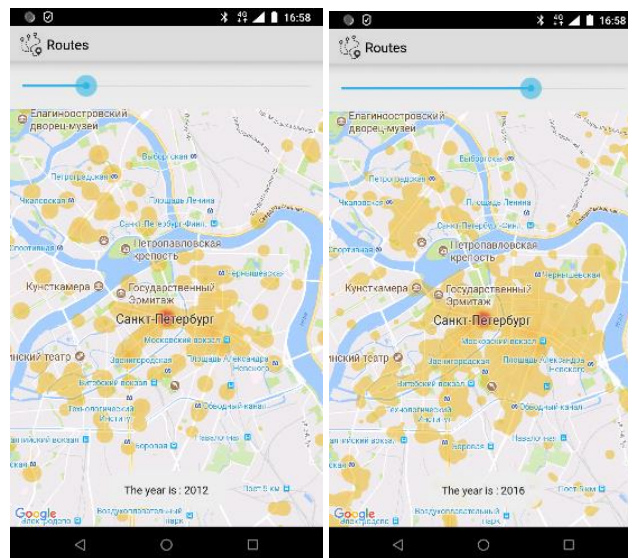


Figure 1 – Situation analysis mode demonstration.

Impact Factor:

ISRA (India) = 1.344	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 0.829	PIHII (Russia) = 0.207	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 4.102	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 2.031	

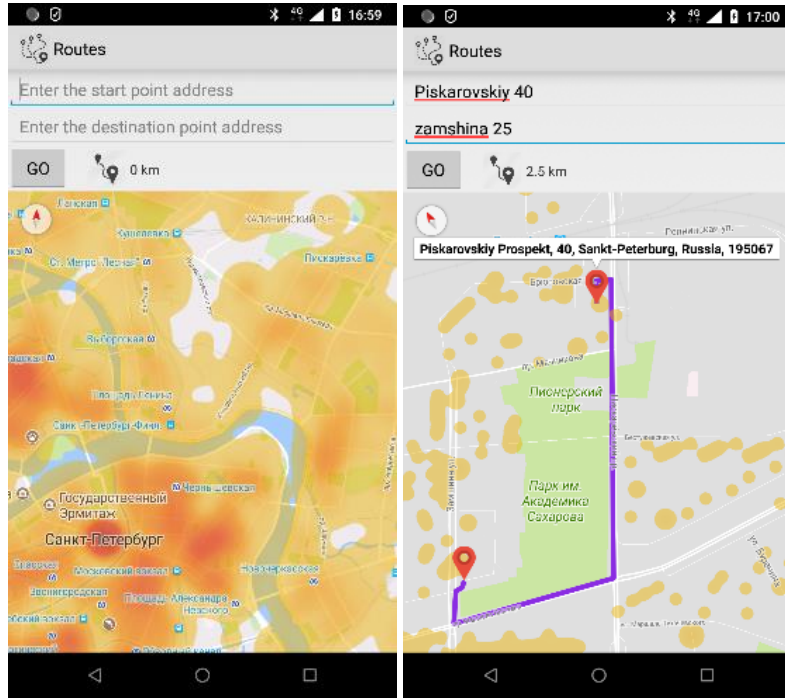


Figure 2 – Pathfinding mode demonstration.

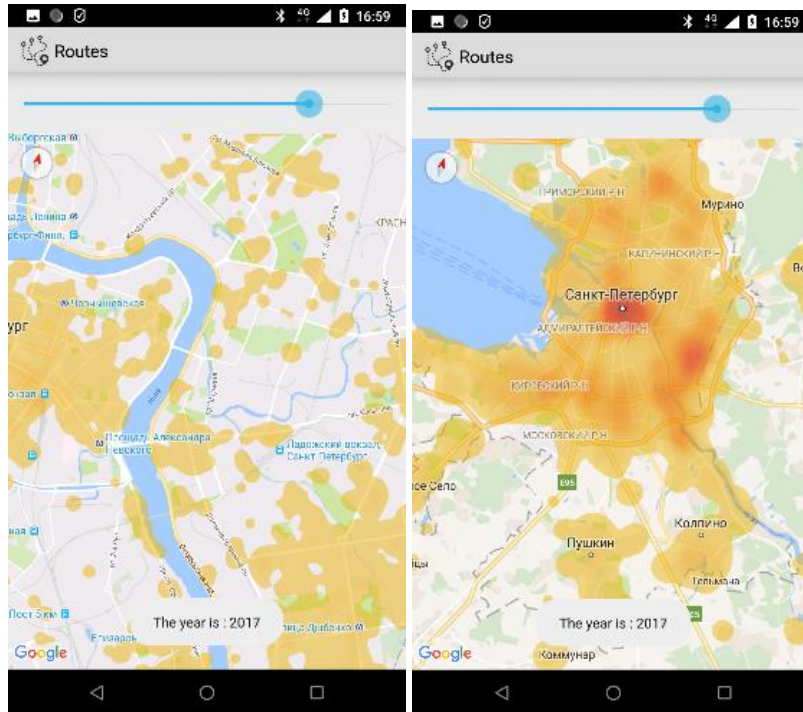


Figure 3 – Heat map behavior demonstration.

Testing

After the application was implemented, the functional testing of the application was carried out. 12 test cases were prepared to test the functionality

of the application. The test cases were launched on 11 devices supporting different versions of the Android operating system. (Table 1)

Impact Factor:

ISRA (India) = 1.344	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 0.829	PIHHI (Russia) = 0.207	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 4.102	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 2.031	

Table 1

Devices that have been used for functional testing.

Device name	Android OS version
Smartphone Huawei Honor 8	Android 7.0.0
Smartphone Nokia 6.1	Android 8.1.0
Smartphone Sony Xperia Z Ultra	Android 5.1.1
Smartphone LG K10 LTE	Android 6.0.0
Tablet Nexus 7 (2013)	Android 6.0.1
Smartphone OnePlus 5	Android 8.1.0
Smartphone OnePlus 3	Android 8.0.0
Smartphone Asus ZenFone C	Android 4.4.2
Smartphone LG G4 H819	Android 4.4.2
Smartphone OnePlus One	Android 7.1.2
Smartphone OnePlus 5T	Android 8.1.0

All tests have been successfully passed on all the above devices, it can be assumed that the application is ready to use.

Efficiency of predictive algorithms

The analysis of the effectiveness of predictive algorithms was carried out in order to determine the trend of year selection, the data of which allow predictive algorithms to construct the best prediction for another year, as well as to establish the most effective algorithms that can be used for similar tasks of geospatial data analysis. To analyze the effectiveness of algorithms, the function of determining the quality of the forecast was written. This function is based on the Euclidean distance. The smaller the value of this function, the better the prediction.

The Euclidean distance is the straight-line distance between two points in Euclidean space. With this distance, Euclidean space becomes a metric space. The associated norm is called the Euclidean norm. A generalized term for the Euclidean norm is the L2 norm or L2 distance. [11]

The analysis to determine the trend of year selection was carried out, the result of the complex prediction of the algorithms was evaluated using the described above function based on Euclidean distance. (Table 2)

It was found that for algorithms training it is better to choose crime data of the year preceding the

year for which the prediction is performed. Predictive algorithms trained on crime data from any year better predict crime rates for the next year than any other. The trend was identified for each training year except 2010. In the case of 2010, this trend may not be apparent since the crime data for 2010 are the smallest and most fragmented of the presented data.

The efficiency analysis of the chosen predictive algorithms was carried out, the results of their work were evaluated using the described above function based on Euclidean distance. (Table 3)

It was found that the linear regression and the support vector machine are the most effective algorithms for solving the problem of the crime level prediction. This result can be explained by the fact that additional regularization for linear regression had been added and the fact that the support vector machine is sensitive to noise in data with which it can handle on its own.

Decision trees and random forest showed the worst efficiency among predictive algorithms, which can be explained by the fact that both methods have a high chance of overfitting. They explain well the examples from the training sample, but they work relatively poorly on the data that did not participate in the training.

Table 2

Results of the analysis to determine the trend of year selection.

Training year			
2010		2011	
Predict. year	Euclidian distance	Predict. year	Euclidian distance
2011	157	2012	92
2012	138	2013	94
2013	110	2014	123
2014	105	2015	205
2015	169	2016	224

Impact Factor:

ISRA (India) = 1.344	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 0.829	PIHII (Russia) = 0.207	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 4.102	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 2.031	

2016	186	2017	235
2017	195		X
Training year			
2012		2013	
Predict. year	Euclidian distance	Predict. year	Euclidian distance
2013	101	2014	71
2014	129	2015	144
2015	218	2016	163
2016	239	2017	175
2017	251		X
Training year			
2014		2015	
Predict. year	Euclidian distance	Predict. year	Euclidian distance
2015	106	2016	33
2016	128	2017	43
2017	140		X
Predict. Year – Predictive year			

Table 3

The efficiency analysis of the chosen predictive algorithms.

Year	Lin. reg.	SVM	Grad. boost	Rand. forest	Dec. trees
2010	70	61	107	144	159
2011	86	81	83	104	116
2012	90	96	100	110	114
2013	60	68	69	79	82
2014	105	104	105	108	110
2015	30	33	31	36	39
2016	21	23	21	24	25

Lin. reg. – Linear regression
SVM – Support vector machines
Grad. Boost – Gradient boosting for regression trees
Rand. forest – Random forest
Dec. trees – Decision trees

Practical significance

The created application can be used by any user, and can also be of interest to people involved in planning the infrastructure of the city, planning the placement of police stations in the city. This application can also be adapted to any other data based, for example, not at the level of criminality in the city.

The application can also be considered as a platform for further research on the possibility of applying data analysis methods to geospatial data. The options of such studies can be the use of neural

networks to data analysis problems, the research about tuning of the gradient boosting algorithm.

Conclusion

A ready-to-use Android application was created. This application allows users build safe routes in St. Petersburg and analyze the crime situation in the city by years. In the future, it is planned to add support of other Russian cities and modes such as building unsafe routes to create routes for the policeman patrols.

Impact Factor:

ISRA (India) = 1.344	SIS (USA) = 0.912	ICV (Poland) = 6.630
ISI (Dubai, UAE) = 0.829	PIHHI (Russia) = 0.207	PIF (India) = 1.940
GIF (Australia) = 0.564	ESJI (KZ) = 4.102	IBI (India) = 4.260
JIF = 1.500	SJIF (Morocco) = 2.031	

References:

- (2018) Crime news of the Ministry of Internal Affairs of St. Petersburg. Available: <https://78.мвд.рф/news/>. (Accessed: 07.06.18)
- (2018) OpenStreetMap data for St. Petersburg with information about the city's amenities and shops. Available: <https://www.openstreetmap.org/#map=11/59.9426/30.2522>. (Accessed: 07.06.18)
- (2018) Open data on St. Petersburg with information on apartment buildings. Available: http://data.gov.spb.ru/opendata/7840013199-passports_houses/. (Accessed: 07.06.18)
- Muller A.C., Guido S. (2017) Introduction to Machine Learning with Python: A Guide for Data Scientists – Sebastopol, CA: O'Reilly Media, Second Edition, 2017.
- (2018) Scikit-learn Documentation. Available: <http://scikit-learn.org/stable/documentation.html>. Accessed: 07.06.18)
- (2018) Python documentation. Available: <https://docs.python.org/3/index.html>. (Accessed: 07.06.18)
- (2018) Google Maps API Documentation. Available: <https://developers.google.com/maps/documentation/>. (Accessed: 07.06.18)
- Yeap E., Uy I. (2014) Marker Clustering and Heatmaps: New Features in the Google Maps Android API Utility Library – Google Maps Platform, 2014.
- (2018) Java Documentation. Available: <http://www.oracle.com/technetwork/java/javase/documentation/index.html>. (Accessed: 07.06.18)
- (2018) Android Development Documentation. Available: <https://developer.android.com>. (Accessed: 07.06.18)
- Deza M.M., Deza E. (2009) Encyclopedia of Distances – New York: Springer, 2009. – p. 94.

