ACTA UNIVERSITATIS AGRICULTURAE ET SILVICULTURAE MENDELIANAE BRUNENSIS

Volume 66 131 Number 5, 2018

SPATIAL DATA DRIVEN EVALUATION OF CITY LOCATIONS

Mikuláš Muroň[1], David Procházka[1]

[1]Department of Informatics, Faculty of Business and Economics, Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

To cite this article: MUROŇ MIKULÁŠ, PROCHÁZKA DAVID. 2018. Spatial Data Driven Evaluation of City Locations. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis,* 66(5): 1301–1306.

To link to this article: https://doi.org/10.11118/actaun201866051301

Abstract

Moving to a new home or setting a new bureau in a new city is always difficult. One does not have knowledge about suitable locations; therefore, people are frequently unpleasantly surprised. High traffic noise, long distance to shops or high criminal activity are just few of many possible disturbing aspects. Certainly, there are many data sources that can help to see some particular aspect of the city life. Nonetheless, it is extremely complex and time-consuming task to browse through large data sets and com-pare provided information. Therefore, we developed a solution that comprises many different data sets that describe the city environment and created set of straightforward indices such as environment, safety, shopping etc. The users just provide the application his/her preferences and the application finds locations that are most suitable for particular cause. The application is presented on the example of the Brno city area.

Keywords: smart city, spatial analysis, GIS

INTRODUCTION

The rising availability of geospatial data opens not only new opportunities, but also rises questions on how to mediate the geospatial data to the public in a meaningful way. One of the possibilities is to use them as a basis for decision-making systems. Municipalities and government offices are collecting and processing an increasing amount of data each year. Published data are, unfortunately, scattered across a variety of data portals, where every institution manages only those from a narrow range of interests. In addition, only raw data are often published, frequently not geocoded or just in a form of a location of a specific phenomenon. This approach satisfies needs of the professional public, but data presented in such a way are not suitable for decision-making, they need to be processed. To extract valuable information, they need to be placed into a broader context. One of the ways to achieve this is by creating indices, which combine available data in a meaningful way and are abstract enough to permit decision-making. Such an approach can answer apparently complicated questions, such as which parts of the city offer a pleasant living environment and why. This methodology has become more popular in recent years (see Jun (2006), Li *et al.*, (2007) and Youtssef *et al.*, (2011)). In this paper we show a possible usage of this approach using an example of evaluations of city locations based on the quality of life.

We propose a methodology, which designs six distinct indicators to map different aspects of the urban environment. Indicators are designed with a respect to user needs and data availability. Compared to other works focused on this topic, we propose to compute indicators in a dynamic manner. The dynamic manner of evaluation enables us to encompass users preferences and needs, and provides results, which are specifically tailored to any user. Utilization of this principle eliminates the problem of presenting just one set of results to a different socio-demographic group with differing needs and requirements.

Literature overview

There have been several studies and projects focused on finding a suitable location within urban areas. Such projects include for example the Opportunity Score (see https://labs.redfin.com), which helps users to find better locations for living according to job possibilities in nearby areas. There are many other, similar projects (e.g. Data2Go, Walk Score etc.), but they are usually more narrowly focused and use only several datasets to obtain the results. Therefore, they evaluate the location just from a specific point of view. In our work, we intend to use a wide variety of data to cover multiple important aspects of city life.

A work by Elmahdi and Afify (2007) presents a GIS Tool for life quality assignment in Egypt. The authors evaluated social, demographic and agronomic attributes and merged them into set of indicators using the Weighted Sum Model. They presented three different variants of indicators to meet specific needs of three different social groups within the population. The weights of each variant of the indicator were set by expert decisions. However, compared to our project, the work of Elmahdi and Afify (2007) does not allow to set weights in a dynamic manner. We argue that the most important input is the personal preference of a particular person, because it's hard to find universal optimum across social groups. Therefore, we propose to enable the user to set their personal preferences and calculate their personal index.

Burian *et al.* (2005) present a model called Urban Planner implemented as an Ersi ArcGIS extension. This tool is capable to detect most suitable areas for spatial development by using multiple-criteria decision analysis methods. The tool is focused on urban planning, rather than an evaluation of the current state, as we do in our project. However, the used methodology is similar to ours and can be used for validation of our approach.

Previously mentioned approach to the construction of indicators has been introduced or commented on by Jensen (2004), Jun (2006), Li *et al.* (2007) and Youtssef *et al.* (2011). In this article, we are primarily going to focus on the description of the evaluation method.

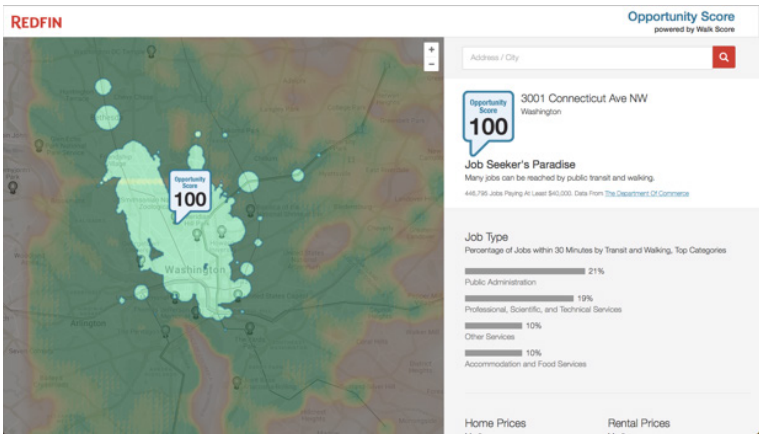
MATERIALS AND METHODS

Data sources

This chapter provides an overview of used data sources. All datasets describe phenomena within the Brno cadastral area. To implement straightforward indicators describing city environment, it is necessary to work with large and heterogeneous data sets. These requirements naturally impose numerous challenges. Available datasets come in different formats (SHP, CSV, GeoJSON etc.) and a lot of pre-processing is needed to make the datasets usable for our purpose. This often requires additional filtering, parsing and geocoding.

Used datasets are provisioned by different sources including government data, publicly available community-created data (e.g. Open Street Maps) and also data from the private sector. The private sector data mostly come from social platforms (e.g. Facebook, Instagram). The essential dataset is RÚIAN (https://www.cuzk.cz/ruian), which provides polygons of city real-estates and is taken as the base layer. The base layer is then used to present calculated values of indicators.

Moreover, the Brno city geoportal (http://data.brno.cz) is among main data sources. It provides a variety of datasets, such as polygons of city parks and green areas, data about traffic, etc. These datasets are intended to be used for the public property management and administration, but we present herein another usage of such data – the use as quality indicators.



1: *Opportunity Score*
source: https://labs.redfin.com/opportunity-score

Regional Office of South Moravia (https://www.kr-jihomoravsky.cz/) also provides useful datasets of predominantly socioeconomic and criminality data and furthermore data describing the quality of environment. Similarly, Ministry of Health and Ministry of Transportation provides data related to noise pollution, location of contaminated places, etc.

Data integration, conversion and synthesis

Datasets need to be processed before they can be integrated into a spatial database. This process constitutes from following steps: a conversion to common coordinate system (WGS-84), geocoding (using Nominatim) and limiting to a boundary box (Brno cadastral area).

Second step is conversion of attributes. Since datasets differ in their granularity (description of phenomena based on different entities e.g. urban areas vs. city districts), we need to convert them into a single type of entity to achieve mutual comparability. We have chosen buildings as the common entity because of their overall orientation to the user-friendly approach. The conversion is done by setting up metrics that convert a value in the original dataset to a new value valid in the context of buildings (e.g. information about a position of a grocery store is converted as a distance from a particular building to the nearest grocery store). We propose several such metrics (e.g. density of phenomenon, distance from POI, sum of attribute etc.) and use them according to the character of a particular attribute. Similar approach has been used also by Sevtsuk (2012).

Indicators

To evaluate locations, different kinds of indicators were proposed. As mentioned, the indicator simplifies a set of complex, spatially distributed variables (e.g. pollution, traffic density etc.) into a single understandable number. Proposed indicators aim to cover most important aspects of the urban environment. Our indicators were designed on the basis of our assumptions related to which attributes of city life are important for different social groups of city inhabitants (students, young families, retired people, etc.). The in-depth description of indicators follows.

Attributes of input datasets were processed in a way, which allows to convert different granularities to a single comparable element. This has been done using multiple techniques, always with regards to the character and structure of the data. Proposed indicators were constructed with respect to available data. For places with a wider variety of available data, additional indicators may be created.

Environment – Ecology indicator

This indicator aims to encompass data related to environmental quality and ecological behavior. It combines available datasets related to the environment and the present integrated insight into the environment quality. The indicator is composed of following attributes:

- level of noise pollution (source: Ministry of Health),
- amount of nearby green areas (source: Brno municipality),
- number of days with exceeded limits of air pollution (source: ČHMU),
- inconveniences caused by the presence of a frequented road (source: Brno municipality),
- distance from a contaminated place (source: Ministry of Health),

To support ecological behavior among citizens, this indicator also includes a distance to a container for recycled waste (source: Brno municipality).

Social status indicator

Social status is multiple-sided phenomenon. For the purpose of this paper, we comprehend it as a relative respect, competence, and deference accorded to people, groups, and organizations in the society (Sauder, 2012). As such, we have used following datasets to estimate its potential value:

- sum of nearby brownfield areas (source: Brno municipality),
- property price (source: Association of Czech real-estate agencies),
- ratio of residential buildings (source: RÚIAN),
- information on whether a building is in a residential area (source: Open Street Map),
- information on whether a building is in a residential street (source: Open Street Map),
- how people feel about a particular locality (source: pocitovemapy.cz).

Safety

Safety is among the key aspects of city life. Therefore, we have dedicated a particular indicator to this purpose, which is composed of yearly density of reported offenses on the street level (source: South Moravian Region Office).

Shopping and services

Availability of public services and shopping possibilities plays an important role in city life. Locations not properly covered by public services experience downfall in overall life quality. This indicator, therefore, encompasses the distance from a particular location to the nearest post office and to the nearest grocery store, as these are among the most frequently used services by residents. To support this indicator, an additional attribute is also included; sum of all nearby commercial areas (source: Open Street Map, Brno municipality).

Fun and Relax

An attractiveness of a location is also indicated by a number of events and cultural opportunities. For that purpose, we have used Facebook data about places and corresponding event attendance. Based on this dataset, this indicator contains an attribute, which maps, how attended are events in a close proximity to a location. Distance and abundance of pubs, bars and restaurants in proximity is used as a supporting attribute (source: Facebook, Open Street Map).

Public transportation

Availability and attainability of public transport also plays an important role in the location quality. This indicator therefore maps the distance to the nearest public tram stop, as well as an average time to get to three other locations in the city (Česká, Mendlovo náměstí, Skácelova) (source: Google Maps Distance Matrix, Open Street Map).

Indicators calculation methods

Attributes are summarized into indicators using two distinct methods. The first method is a simple multiple criteria decision-making method – Weighted Sum Model. The value of an indicator is calculated as a sum of attributes in the corresponding indicator. Weights are set to each of the attributes based on the assigned importance of the attribute in the indicator.

We need to ensure that this index will be relevant for users with different needs. To achieve this, we implement weight setting mechanism, which enables setting a weight to each individual indicator. Weight value is determined by users themselves according to their needs.

We also implement a second method based on the Principal Components Analysis, which is used for reduction of dimensions. This method has been widely used in similar cases by e.g. Jun (2006). To allow weightings of attributes, we use a slightly modified version of this method called the Weighted Principal Component Analysis introduced by Delchambre (2015).

Methods for weights settings

To allow users to state their preferences, we implement weights setting technique based on an assignment of a finite number of points to each indicator. These points evaluations are then recalculated to weights on the percent basis.

We have also tested the Analytical Hierarchy Process for weight assignment. This method needs a direct comparison of a set of tuples. This set can be quite large and the overall process is very time consuming. During the testing phase, we have decided to remove this approach from the web application after the feedback from users.

RESULTS

Calculated indicators are presented via an interactive web application available at http://brno.ml. This application presents a calculated evaluation for each indicator and also allows users to perform an on-demand calculation based on their preferences.

The application consists of two main parts. Web interface for presenting the results uses Leaflet and Carto.js. Server side part provides API written in Python for on-demand calculation and a CartoDB platform for serving map tiles with results. Data are stored in PostGIS spatial database.

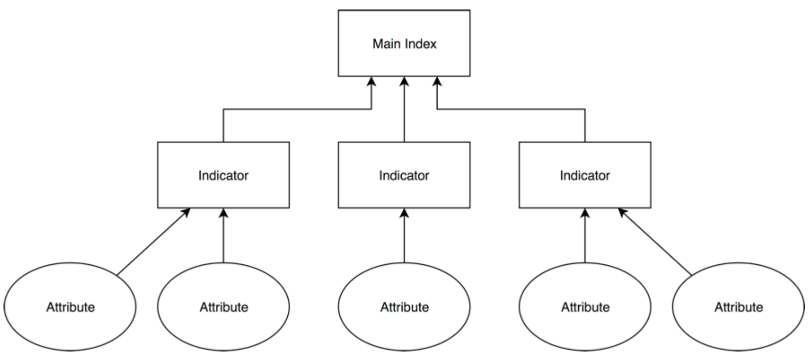
When a user enters the app, main index evaluation with the default values is presented. The user may enter their preferences into the system by assigning available points to each indicator. The amount of points is limited and user is therefore forced to decide and formulate priorities. Based on this assignment, on-demand main index calculation is performed. To allow backtracking of presented results, it's possible to see attributes and their corresponding values for a selected real-estate in the detailed overview (see Fig. 3).

DISCUSSION

Herein introduced application can be used by residents to support their decision-making, when choosing a new place of residence. Presented methodology is universal enough to be adapted to other cities, as well, without many modifications, but it is always limited by available data.

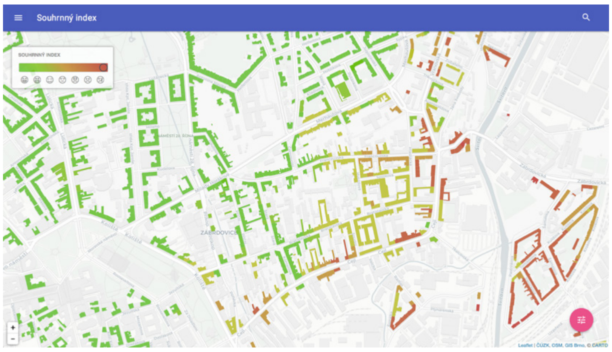
Accuracy evaluation

We used several methods to evaluate accuracy of results. The key method was a survey among Brno city residents. This survey compared values of indicators calculated in our application (on the basis of particular user preferences) with an evaluation gathered by the city resident. Results of this survey



2: *Hierarchy of indicators*

confirmed relevance for most indicators (see Tab. I). Certainly, there were disparities in the evaluation in some cases, but almost exclusively in the case of attributes that are highly subjective (level of services in the area etc.). There is no precise value, which would describe, when a shop or restaurant is "too distant" or how many shops are necessary for "good" public services.



3: *Application shows calculated main index evaluation on a real-estate base layer.*

4: *Example of results for Environment – Ecology indicator. Note changes in values around a frequented road vs. a park area.*

5: *Example of results for Shopping and services indicator. Note changes in values between the city center and periphery.*

I: *Results of accuracy evaluation. The value represents a difference between the calculated index and user opinion about a locality.*

Indicator	Average	Residence	Locality 1	Locality 2	Locality 3
Environment	1.34	1.34	1.55	1.24	1.25
Social status	1.93	2.03	1.92	1.2	2.58
Shopping	2.88	2,98	2.18	3.12	3.17
Safety	1.79	2.12	1.78	0.82	2.45
Fun and Relax	2.70	2.77	2.14	2.94	3.02
Transport	1.89	1.98	2.17	1.45	1.98
Average	2.08	2.20	1.95	1.80	2.40

CONCLUSION

We argue that the proposed methodology and implementation in this paper is very general; hence, it can be used as a basis for similar applications for any city. However, it is necessary to find suitable datasets. Especially the low quality of the source data can cause misleading results. We must emphasize that such an application (and similar ones) has only a limited value for inhabitants of a particular city. Knowledge accumulated over years cannot be encompassed in a simplified number. However, it can bring an interesting insight for people that are not familiar with a particular city.

REFERENCES

BURIAN, J. BRUS, J. and STASTNY, S. 2015. Urban Planner–model for land use suitability assessment. *New Developments in Environmental Science and Geoscience*, 19: 20–29.

DELCHAMBRE, L. 2015. Weighted principal component analysis: a weighted covariance eigendecomposition approach. *Monthly Notices of the Royal Astronomical Society*, 446(4): 3545–3555.

ELMAHDI, A. and AFIFY, A. 2007. Development of a GIS tool for qualitative assessment of the Egyptian's quality of life. *The Environmentalist*, 27(1): 183–194.

JENSEN, R. et al. 2004. Using remote sensing and geographic information systems to study urban quality of life and urban forest amenities. *Ecology and Society*, 9(5): 5.

JUN, B.-W. 2006. Urban Quality of Life Assessment Using Satellite Image and Socioeconomic Data in GIS. *Korean Journal of Remote Sensing*, 22(5): 325–335.

LI, G. and WENG, Q. 2007. Measuring the quality of life in city of Indianapolis by integration of remote sensing and census data. *International Journal of Remote Sensing*, 28(2): 249–267.

SAUDER, M., LYNN, F. and PODOLNY, J. M. 2012. Status: Insights from organizational sociology. *Annual Review of Sociology*, 38: 267–283.

SEVTSUK, A. and MEKONNEN, M. 2012. Urban network analysis: a new toolbox for measuring city form in ArcGIS. In: *Proceedings of the 2012 Symposium on Simulation for Architecture and Urban Design*. Society for Computer Simulation International.

YOUSSEF, A. M., PRADHAN, B. and TARABEES, E. 2011. Integrated evaluation of urban development suitability based on remote sensing and GIS techniques: contribution from the analytic hierarchy process. *Arabian Journal of Geosciences*, 4(3–4): 463–473.

Contact information

Mikuláš Muroň: mikulas.muron@mendelu.cz
David Procházka: david.prochazka@mendelu.cz