2018 IEEE International Conference on Smart Computing

Demo: Bostonhood: a multi-criteria platform for ranking city neighborhoods

Despoina Kazepidou
Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki, Greece
dkazepid@csd.auth.gr

Maria Rousi
Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki, Greece
mariarousi@csd.auth.gr

Vasiliki Gkatziaki
Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki, Greece
vasgat@gmail.com

Athena Vakali
Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki, Greece
avakali@csd.auth.gr

***Abstract*—Recent advancements in information and communication technologies resulted an increase on data produced by the cities. These data are either produced on social media networks or collected by public services and are usually enriched with useful geospatial information. This influx of information coupled with the urge of urban designers to make cities smarter, led to the development of intelligent applications designed to facilitate the everyday lives of citizens. In this demo we present an online platform, Bostonhood, which is responsible for neighborhood recommendation and sustainability assessment. Data came from the official open data hub for the city of Boston, the location-based social network Foursquare and the Airbnb platform. The methodology that we followed combines two different Multi-Criteria Decision Analysis algorithms (COPRAS and TOPSIS) and creates a mutual ranking using the Borda count method. This platform is useful not only to citizens and tourists who want to find the best neighborhood to live in according to selected criteria but also to urban designers who want to improve Boston's sustainability.**

Keywords—Multi-criteria decision analysis, Data mining, Smart city, Neighborhood sustainability

I. Introduction

Cities produce daily a vast amount of data. These data are produced by various sensors placed in different locations in the city, observations of urban designers, a variety of stakeholders, citizens using tools which are installed on their mobile phones, public documents and social networks. These data contain a large amount of useful geospatial information. Also, this influx of information coupled with the urge of urban designers to make cities smarter, leads to the development of smart applications designed to facilitate the everyday lives of citizens (e.g. DynamiCITY [1], CityPulse [2]).

The aim of this demo paper is to demonstrate how we could leverage the new possibilities and information arising from the modernization of cities to solve specific problems. In detail, this work addresses the problem of discovering neighborhoods of residence that meet the needs of both residents and tourists of a city. These needs as well as the criteria for choosing a residential area have changed and this is largely due to the problem of urbanization. The criteria are not only economic but also social and environmental [3][4]. Residents want to live in sustainable areas of residence, while tourists, having no information on the city they visit, want to discover the closest solution to their needs and desires. Consequently, it is necessary to assess the sustainability of the different areas of a city, which will also assist in identifying and solving its problems by the competent authorities.

Bostonhood is an online platform, which is responsible for neighborhood recommendation and sustainability assessment for the city of Boston. Nevertheless, the methodology is general and not limited to the Boston area. By collecting the appropriate data, it could be used in other cities or regions by adding or removing some of the criteria. The criteria used in this implementation, in order to achieve sustainability evaluation, are Environment, Education, Safety, Arts & Entertainment, Food & Nightlife, Shops & Services, Arts & Culture. Our platform exploits open data from different sources: the official open data hub for the city of Boston, the location-based social network Foursquare and the Airbnb platform. Apps of similar content usually use open data from official government sites rather than social networks. The methodology that we follow combines two different Multi-Criteria Decision Analysis algorithms (COPRAS and TOPSIS) and creates a mutual ranking using the Borda count method. As both MCDM algorithms require weights, Rank Sum method is used for their estimation. Finally, statistics are calculated to check the efficiency of the algorithms.

This platform can be used by many different recipients: inhabitants, tourists and urban designers. Inhabitants can find the neighborhood closest to the criteria that the ideal residential neighborhood should have, tourists can find the neighborhood which is closer to the criteria that the ideal neighborhood for accommodation should have and receive houses suggestions in this region and urban designers can evaluate neighborhoods based on the criteria selection, so that they would know in which criteria each neighborhood lacks and take care to improve them by taking appropriate decisions. Bostonhood platform aims to provide guidelines on decision-making issues which recipients encounter and improve sustainability of Boston neighborhoods by improving specific aspects (environmental, social, economic).

978-1-5386-4705-9/18/$31.00 ©2018 IEEE
DOI 10.1109/SMARTCOMP.2018.00060

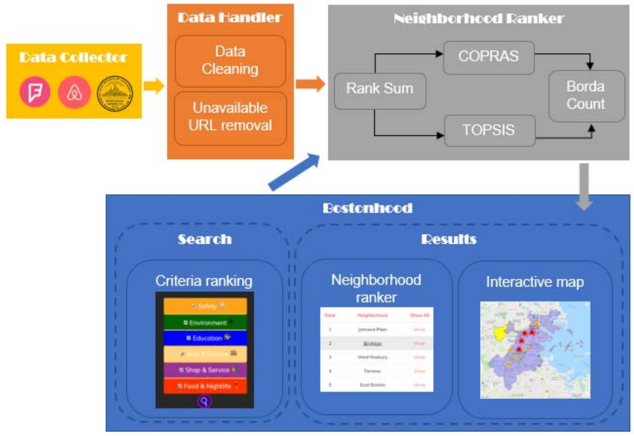

Fig. 1. Bostonhood platform architecture.

II. ARCHITECTURE

Figure 1 presents the architecture of the Bostonhood platform. It shows the basic methodology we followed to evaluate neighborhood sustainability. In detail, Bostonhood platform consists of two main components:

- A backend component, which is responsible for the collection and preprocessing of the data and also for the implementation of the MCDM algorithms. The Data Collector module collects data from Foursquare, Airbnb and the official open data hub for the city of Boston[1]. The Data Handler module is responsible for data cleaning and preprocessing. In detail, data with no geospatial information available and data from Airbnb with dead URLs are removed. Additionally, all the non-ASCII characters from the dataset are deleted. The next module, Neighborhood Ranker, includes the main processes of the methodology. Firstly, a sustainability profile is created for each neighborhood by using the collected data. Consequently, the weights of the criteria are calculated according to the Rank Sum method, based on the user's preferences and they are used as input to the MCDM algorithms (TOPSIS, COPRAS). The results of these two algorithms, i.e. the ranking of Boston's neighborhoods according to sustainability, are combined using Borda count method and the final results are fed into the frontend component of the platform.
- A frontend component, which is responsible for the presentation of the results. An interactive map of the Boston area is used to visualize the results, i.e. the ranking of the neighborhoods from the most to the least sustainable, according to user's preferences and some statistics regarding the neighborhoods.

2.1 Rank Sum

Rank Sum method is leveraged to set weights to criteria that depict their significance. These values are used in the Multi-Criteria Decision-Making methods which are described in the following section. In Rank Sum method, the criteria are ordered, and their values are defined using the following formula, where *i* corresponds to each criterion rank and *n* to the sum of the ranks [5]:

$$w_i(RS)=\frac{n+1-i}{n(n+1)/2},\ i=1,\ldots,n$$

2.2 Multi-Criteria Decision Making algorithms

In this section we describe two of the most popular Multi-Criteria Decision-Making algorithms: COmplex PRoportional ASsessment (COPRAS) and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS). Both of the algorithms belong to the Multi-Attribute Decision Making (MADM) techniques, because alternatives are predefined [6]. Several studies have been conducted [3][4] that use these algorithms to evaluate sustainability of an area when there are multiple criteria. Their calculations may vary, but both algorithms require decision matrix creation, normalization and weighting. Each algorithm results a ranked list of the alternatives in descending order according to the chosen criteria.

2.3 Borda count

Borda count [7] is a consensus-based election method which is used when different rankings must be combined. In MCDM problems, it has been used to join the results of two or more different MDCM algorithms [7]. There are many variations of this method where each alternative is graded differently. We used the Borda's system (starting at 1). Using this method, we combined the different rankings of Boston's neighborhoods from COPRAS and TOPSIS to find the most widely accepted solution.

III. DEMONSTRATION SCENARIO

Our datasets came from three different sources, the official open hub for the city of Boston, the location-based social network Foursquare and the Airbnb platform. The Airbnb dataset was found in Kaggle[2] platform. All data include

[1] https://data.boston.gov/

[2] https://www.kaggle.com/

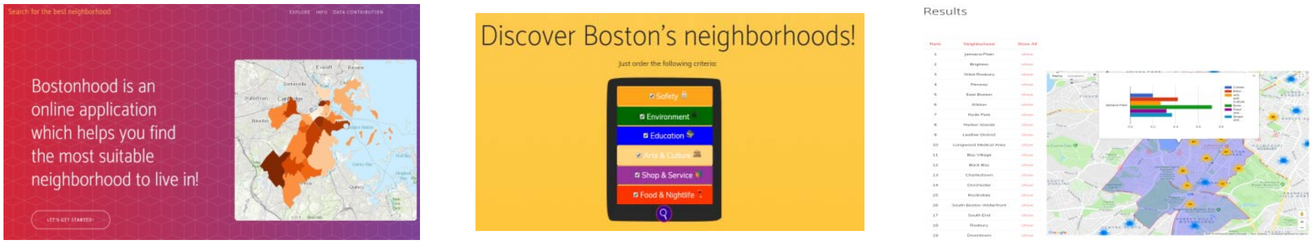

Fig. 2. Bostonhood platform demonstration.

geospatial information and are up to date until 2017. We distributed them, according to their location, to the 26 neighborhoods of Boston, whose boundaries are presented as polygons in a GeoJson file which comes from the official open data hub of Boston. Foursquare is a location-based social network which we harnessed to collect venues from certain categories (e.g. Food & Nightlife) for the 26 neighborhoods of Boston. In detail, we managed to collect from Foursquare 7,911 venues. The Airbnb dataset contained the available houses in Boston neighborhoods. After removing the houses which are not available anymore, the dataset comprised of 2,216 houses. Finally, the data from the official site of Boston contained Points of Interest or Incidents for the different criteria. In total, the dataset contains 422,335 data points. By using these data, we created a sustainability profile for each neighborhood of Boston which contains the number of Points of Interest or Incidents for each criterion.

A screen shot of Bostonhood platform is presented in Figure 2. Firstly, there is a short description of the platform. In this screen the user has 3 options: 1) click "Explore" or "Let's get Started" to go to the next page and order the criteria 2) click "Info" to go to the 3rd page and read some information about the criteria that are used in this platform 3) click "Data contribution" to go to the last page and learn more about the sources that were used for data collection and navigate to the sites of the sources.

To begin the demonstration, the user needs to select and order the criteria according to his/her preferences. When they are ready, they click the search button and a new screen appears provided that at least one criterion has been chosen. This new screen shows the results of the Borda count method using a list and an interactive map. The list contains the names of the neighborhoods from the most suitable to the least suitable, according to the chosen criteria. The user can click on the name of the neighborhood or show button to view its location on the map. In addition, statistics for each neighborhood are shown if the user clicks on a region on a map. Their values have been calculated using the following formula, where *c* stands for the criterion, *n* for neighborhood, *m* for the neighborhood which has max number of Points of Interest or incidents for criterion *c* and *Score(i,j)* stands for the score of neighborhood *j* in criterion *i*, i.e. the number of POIS or incidents for neighborhood j and criterion i:

$$\frac{\text{Score}(c, n)}{\max[\text{Score}(c,m)]}, \quad c=1,2,\ldots,6, \;\; n,m=1,2,\ldots,26$$

So, the values of the statistics get values from 0 to 1. Lastly, there are markers on the map with numbers showing the available Airbnb houses and their exact addresses. This whole procedure is executed dynamically when the user orders the criteria and clicks the search button.

IV. Conclusion

Bostonhood is an online platform which proposes the ideal neighborhood to live in within Boston city by evaluating each neighborhood's sustainability according to the criteria that the user has selected. The results of our experiments show that in terms of sustainability, the best neighborhood is Dorchester, while the worst seems to be Chinatown. In addition, Kendall's Tau correlation coefficient show that in most cases there is a strong correlation between COPRAS and TOPSIS methodologies. Finally, from the z-values we conclude that the results of the methods are statistically significant at 93.56% for a 5% significance level.

Acknowledgment

This work has been supported by the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No. 780121.

References

[1] V. Gkatziaki, M. Giatsoglou, D. Chatzakou, and A. Vakali, "DynamiCITY: Revealing city dynamics from citizens social media broadcasts," Information Systems, vol. 71, pp. 90–102, Nov. 2017.

[2] M. Giatsoglou, D. Chatzakou, V. Gkatziaki, A. Vakali, and L. Anthopoulos, "CityPulse: A Platform Prototype for Smart City Social Data Mining," Journal of the Knowledge Economy, vol. 7, no. 2, pp. 344–372, 2016.

[3] E. Mulliner, K. Smallbone, and V. Maliene, "An assessment of sustainable housing affordability using a multiple criteria decision making method," Omega, vol. 41, no. 2, pp. 270–279, Apr. 2013.

[4] M. Taghvaei, H. R. Varesi, and M. Narimani, "An Analysis on Effect of Urban Development Plans on Realization of Sustainable Development of Metropolis of Isfahan," Modern Applied Science, 2016.

[5] M. B. Barfod and S. Leleur, "Multi-criteria decision analysis for use in transport decision making," DTU Transport Compendium Series part 2, pp. 1-76, 2014.

[6] R. Z. Farahani, M. SteadieSeifi, and N. Asgari, "Multiple criteria facility location problems: A survey," Applied Mathematical Modelling, vol. 34, pp. 1689–1709, 2010.

[7] R. Volvačiovas, Z. Turskis, D. Aviža, and R. Mikštienė, "Multi-attribute Selection of Public Buildings Retrofits Strategy," Procedia Eng., vol. 57, no. Supplement C, pp. 1236–1241, Jan. 2013.