Land Use Policy 54 (2016) 253–263

Contents lists available at ScienceDirect

Land Use Policy

journal homepage: www.elsevier.com/locate/landusepol

A data based model as a metropolitan management tool: The Bogotá-Sabana region case study in Colombia

Daniel Felipe Del Busto Pinzón[a], Fabio Teodoro de Souza[b,*]

[a] *Low Carbon Resilient Development Program Colombia—USAID/USFS, Carrera 18 No. 86A-14, Bogotá 110221, Colombia*
[b] *Pontifical Catholic University of Paraná (PUCPR), Postgraduate Program in Urban Management (PPGTU), Polytechnic School and School of Architecture and Design, Rua Imaculada Conceição—1155-Prado Velho, Po Box: 16210, 80215-90, Curitiba, Paraná, Brazil*

ARTICLE INFO

Article history:
Received 16 September 2015
Received in revised form
13 November 2015
Accepted 18 February 2016

Keywords:
Data mining
Urban management
Metropolitan region
Clustering
Association rules

ABSTRACT

Data mining techniques are useful tools for knowledge extraction in large databases. They are successfully applied in disciplines such as urban management and geography. The use of these techniques is suitable for the study of complex urban dynamics of post-industrial cities. The objective of this research is to empirically apply data mining techniques as a tool for metropolitan management. This paper presents a methodology developed to elucidate the dynamics of urban growth and its relationship with socioeconomic patterns in the Bogotá-Sabana region of Colombia. To analyze the evolution of growth in two time frames, various techniques are applied, including correlation, clustering, regionalization, and others. The results enable interesting patterns to be identified in urban development. A key aspect of these patterns is the high level of segregation and inequity that is apparent in the case study. Finally, the application opportunities of these tools in urban management are discussed.

1. Introduction

Since the 1960s, data mining techniques (DMTs) have been developed as artificial intelligence methods that are applied for the extraction of valuable knowledge from large volumes of data (Han et al., 2011). The explosive growth of database technology, as well as the advancement of information and communication technologies (ICTs), has promoted the development of such tools for intelligent information processing (Liao et al., 2012). DMTs enable the analysis of databases that are large (e.g., millions of observations), highly dimensional (i.e., hundreds of variables), and highly complex (e.g., heterogeneous data sources, space-time dynamics, multivariate relations) (Mennis and Guo, 2009). For these reasons, DMTs are successfully applied in various disciplines, including geography, urban management, and transportation.

At the same time, the ICT infrastructure has enabled the consolidation of the globalization process. New logics of spatial organization have materialized in cities in accordance with the principles of this process (Méndez, 2012). The largest cities of Latin America have ensured their positions as the main nodes of coordination between national and international economies. This fact has stimulated the expansion of urban areas beyond administrative boundaries, increased the tension in peri-urban areas, and deepened social segregation (Avila Sánchez, 2009; De Mattos, 2010). According to statistics, between 1970 and 2011, the urban population in Latin America grew at a rate of 2.59%, which was in contrast to the 0.03% increase of the rural population. These data substantiate the fact that the degree of urbanization in Latin America (79.1%) is currently higher than that of the European continent (72.9%) (United Nations, 2012).

Owing to the urban challenges that the cities of Latin America present, the main objective of this research is to evaluate the potential application of DMTs as innovative tools for metropolitan management. As a case study, the research focuses on the urban growth of Bogotá and four surrounding cities: Funza, Mosquera, Chía, and Soacha. These municipalities have the highest values in the metropolis index in relation to Bogotá (Alfonso, 2010). The analysis is performed based on the following techniques: the hierarchical tree, principal component analysis, *k*-means regionalization, and association rules (Han et al., 2011; de Souza, 2004).

The paper first highlights the need to include innovative strategies for city management. Next, the contextualization of the case study is presented. This is followed by a description of the methodology according to the three basic steps of any data mining project: data collection, preparation of the databases, and modeling. The results of the empirical application are then described, and a discus-

* Corresponding author.
E-mail addresses: felipe.delbusto@gmail.com (D.F. Del Busto Pinzón), fabio.teodoro@pucpr.br (F.T. de Souza).

http://dx.doi.org/10.1016/j.landusepol.2016.02.019

sion of the opportunities of these tools within urban management concludes the paper.

2. Data mining in metropolitan management

Since the middle of the previous century, globalization has been considered the paradigm of the transformation and understanding of various contemporary realities (Castells, 1999; Santos, 1994). The restructuring of the capital accumulation system, supported by the ICT infrastructure, has socially and spatially reconfigured the metropolises of the world, including the cities of Latin America (De Mattos, 2010; Méndez, 2012). Currently, the post-industrial metropolis is characterized by highly complex territories. On one hand, there are peri-urban areas of rapid growth and transformation that defy the traditional rural-urban dichotomy (Allen, 2003; Avila Sánchez, 2009; Dematteis, 1998). On the other hand, there are polycentric urban networks composed of multiple local entities that maintain relationships and cross-scale flows (Méndez, 2012; Precedo Ledo, 1996).

In this sense, the development of ICTs has had an enormous impact on various aspects of societies; moreover, its influence has expanded in the study of urban management. An increase in the ability to generate, store, and process data, and performance improvement in using the knowledge generated via the same technology, have created a growing circuit of technological innovation (Castells, 1999). As a consequence, in the urban context, multiple terms have been proposed—the informational city, learning city, intelligent city, knowledge-based city, etc. These terms are aimed at explaining the key criteria relating to the generation and use of information/knowledge with increasing innovative responses for the management of cities (Méndez and Sánchez Moral, 2010).

Within this circuit of innovation, the systematic evolution of database technology, which has enabled the emergence of data mining and the process of knowledge discovery, has been highlighted (Han et al., 2011). Thus, since 1960, DMTs have formed a set of methods that includes applied artificial intelligence, which takes into consideration the tasks of generalization, characterization, classification, association, grouping, regionalization, forecasting, etc. (Liao et al., 2012). The application of DMTs covers a multidisciplinary field of research, including geography, urban planning and management, among many others. This multidisciplinary application is due to the difficulty of analyzing large geographic scales and the complexity of territorial processes, as well as the continued efforts of government agencies, scientific projects, and private entities in generating information. It is also due to the development of modern techniques of collecting data, such as geographic positioning systems (GPS), high-resolution remote sensing, automatic location detection systems, and geographic locations voluntarily provided by Internet users (Mennis and Guo, 2009).

In specialized literature, DMTs are used to analyze various urban phenomena. Examples of research conducted by applying DMTs include the study of physical and meteorological events, calculation of the metropolization level, knowledge-based measurement of the urban development stage, and analysis of the impact of architecture on urban vitality. The assessment of the impact of transportation systems and the modeling of the attributes of daily travel are other examples (Bocarejo et al., 2013; De Cos Guerra, 2007; Kusakabe and Asakura, 2014; Méndez and Sánchez Moral, 2010; Netto et al., 2012; de Souza, 2004). Nevertheless, data mining remains an emerging research area that requires the formulation of theory and methodology that leverages its utilization (Mennis and Guo, 2009).

3. Case study context

Although no legally defined metropolitan area exists, the metropolitan area of influence in Bogotá expands to more than 16 municipalities of the Province of Cundinamarca (Carvajal, 2012; Wessels et al., 2012). This lack of metropolitan arrangements has historical roots that generate uncertainty on the part of local governments in relation to the central power of Bogotá (Frey, 2013; Pecha et al., 2011). However, the metropolitan influence is evident in the high degree of structural mobility of the population (changing residences) as well as in the high degree of everyday mobility (work and study). This is especially the case in the four neighboring cities of Bogotá that have the highest values of the metropolis indexes: Chía, Funza, Mosquera, and Soacha (Alfonso, 2010).

As the capital of Colombia, Bogotá is the demographic and economic center of the country. The city concentrates more than 15% of the population of Colombia in 355 km^2 of urban area, with approximately 7,500,000 inhabitants (DANE, 2010). With a population density of 20,500 inhabitants per square kilometer, Bogotá is one of the densest cities compared to other cities and metropolitan areas of the world (SDP, 2011a). In economic terms, Bogotá represents nearly 25% of the national GDP, with production mainly based on the development of advanced services (SDP, 2011b). The other municipalities considered within the metropolitan cluster of Bogotá, 16 neighboring municipalities, include more than 1,500,000 inhabitants (DANE, 2010) and account for approximately 3% of the national GDP (SDP, 2011b).

In social terms, Bogotá has a very intense level of segregation (PNUD, 2008) that spatially manifests as follows. The southern and western suburbs are characterized by the presence of low-income families, with a low level of formal job opportunities and the highest population density rates. On the other hand, in the north and east, where the expanded city center is located, high-income families, lower population densities, and the majority of formal employment exist (SDP, 2011a). As Fig. 1 shows, this division is perceived in the spatial distribution of the legal social stratification that enables the differentiation of public service tasks. Strata 1, 2, and 3 are characterized by a population with a low payment capacity; this population receives subsidies for the payment of services. Strata 5 and 6 have a population with a higher level of income; this population pays extra to subsidize. Finally, the Stratum 4 population does not receive subsidies or pay extra costs.

According to urban land values, these trends seem to expand to surrounding municipalities. To the south, the price of land in Soacha only surpasses slightly more than a quarter of the amount corresponding to the municipality to the north, Chía. On the other hand, the municipalities to the west of the cities Mosquera and Funza retain their values at about half of this reference value (SDP, 2011b).

For urban management purposes, Bogotá was divided into 112 territorial units, called *Unidades de Planeación Zonal* (UPZ). Along with the urban areas of Soacha, Chía, Funza, and Mosquera, the UPZ are used as units of analysis in this research. This study area is the most dynamic migratory basin in the country, considering that both Bogotá and the Sabana region have positive migration balances compared to the population movement in the *Departmento of Cundinamarca* and in the country. This dynamic is so significant that the municipality of Soacha has greater urban growth than the other cities of Colombia (Alfonso, 2010).

The demographic structure of the study area shown in Fig. 2 is based on the population density and growth rate for 2011. For population density, Bogotá includes an area in the south that is increasingly saturated, especially those UPZs that are more peripheral and close to Soacha, where the limit of 280 inhabitants/ha is exceeded. A similar situation occurs in the periphery of the northwest. In contrast, the areas around the corridors connecting the expanded center with the far north and far west hardly surpass 200 inhabitants/ha.

The growth rate between 2005 and 2011 has two main interpretations. First, a distribution of the highest growth rates exists