Discovering Patterns of Urban Development

Ilija Ilievski[1], Sonja Gievska[1], Ognen Marina[2]

[1]Faculty of Computer Science and Engineering, University of "Ss. Cyril and Methodius", Skopje, Macedonia
ilija139@gmail.com, sonja.gievska@finki.ukim.mk
[2]Faculty of Architecture, University of "Ss. Cyril and Methodius", Skopje, Macedonia
ognen.marina@arh.ukim.edu.mk

Abstract. The goal of this research is to develop a tool that employs intelligent technologies to capture the patterns of urban change driven by a diverse set of context factors. Data mining provides opportunities, which complement and extend the knowledge previously obtained with other approaches. A series of exploratory case studies are in place to investigate the modeling and predictive capabilities of the tool. A number of simulations have revealed distinctive local patterns of urban change in the city of Skopje, shaped by local urban spatial and institutional structures. This study shows the importance of intelligent technologies in the interpretation of the historical evidence of urban development.

Keywords. Urban dynamics, urban development prediction, knowledge-based approach, data mining.

1 Introduction

The idea of a city as an urban space that is sustained by human connections requires suitable models that recognize the sound interplay between physical and phenomenological aspects of the urban system. Urban systems are complex systems, very difficult to delimitate, evolving through space and time. If the general structures of urban systems are complex, the processes and behaviors that can be observed are far from being entirely stochastic [1]. Our knowledge of urban phenomenon might have been deepened on many levels, though its modeling and computational representation is still an open problem [2].

Urban models are envisioned as tools that can help decision-makers anticipate, plan and design cities that promote well-being, prosperity and sustainability. Designing livable and sustainable cities is a shared vision of both, decision makers and people who live there. Experimentation with predictive urban models are expected to lead the way to a better understanding of the general processes and conditions at play in the urban world.

Past analytical methods for studying urban dynamics were greatly limited, mainly because of the kind of data and research tools that were available. The interdisciplinary nature of the problem under consideration requires a diverse set of researchers

V. Trajkovik, A.Mishev (Editors): ICT Innovations 2013, Web Proceedings, ISSN 1857-7288

(architects, computer scientists, sociologists) and efforts that exceed the expertise of any individual field. The new digital, pervasive and intelligent technologies have brought more information to consider, though the process of their utilization is far from straightforward. Thanks to the achievements in domains such as multimedia, knowledge management, machine learning and data mining technologies, computer support of urban planning received new resources and can move to new directions, e.g. creating intelligent simulation and modeling tools and systems.

There are a number of approaches for formalizing and testing hypothesis about different aspects of urban dynamics, finite state models and agent-based paradigm have established themselves as most prominent ones. Models of urban dynamics also vary in the level of detail and abstraction as well as the focus or aspects of development under investigation, from representation of urban entities (e.g., residents, developers and government) when exploring the evolution of spatial relationships over time [3] to studies of urban sprawl determined by the interactions of environmental and demographic factors [4] to a land-use model representing the relationships between the landscape characteristics and the preferences and behaviors of various actors [5].

This research attempts to test a series of hypotheses on how the mechanisms of social and economic stratification have manifested in urban space and whether population dynamics has reconfigured the spatiality of the city landscape. This paper reflects upon the challenges surrounding the efforts in recognizing and interpreting the patterns of urban change. Our efforts are directed towards correlating real-world emergent patterns of urban change to contextual knowledge and incorporating them into model's predicting capabilities. Suitability of an extensive set of machine learning algorithms for simulation and prediction of urban development is investigated. Our long-term goal has been to layout a foundation in terms of a knowledge base and a tool that will accommodate future exploration of different research scenarios related to urban dynamics.

2 Integrated Urban Knowledge

An integrated knowledge base has been proposed as a basis for urban models, a container of a variety of traceable information needed for semantic description of the urban context. Our research has two interconnected objectives: (1) to explore the feasibility of creating urban knowledge base and a tool to support the construction of models of urban dynamics; and (2) to demonstrate the usefulness of this tool in terms of exploratory scenario-based case studies.

The knowledge base (Fig. 1) should provide a means for integrating and interconnecting heterogeneous data formats such as urban maps, photographs, cadastre data and various unstructured data (A), as well as census data, empirical studies and social surveys (B). This effort needs access to data, solicited and gathered by experts in various fields (e.g. architects, city planners, local government, social science experts) with various solicitation and analytical methods. Semantic heterogeneity, terminology differences, inconsistency, redundant data and interoperability are some of the problems to be encountered.

The new directions in information technologies aimed at pervasiveness and intelligence have increased the amount of raw data collection with a potential to increase our knowledge of different aspects of social urban life. The employment of a number of tools and intelligent techniques could support the process of capturing and visualizing the observable manifestation of behavior trends and patterns i.e. the urban dynamics (C). Extracting qualitative knowledge from large quantities of data is just the beginning of our search for meaning and plausible explanation of urban dynamics. New platforms that combine urban informatics with the more diverse urban-related knowledge are yet to be developed and deployed (D). This research is an attempt in support of those efforts.

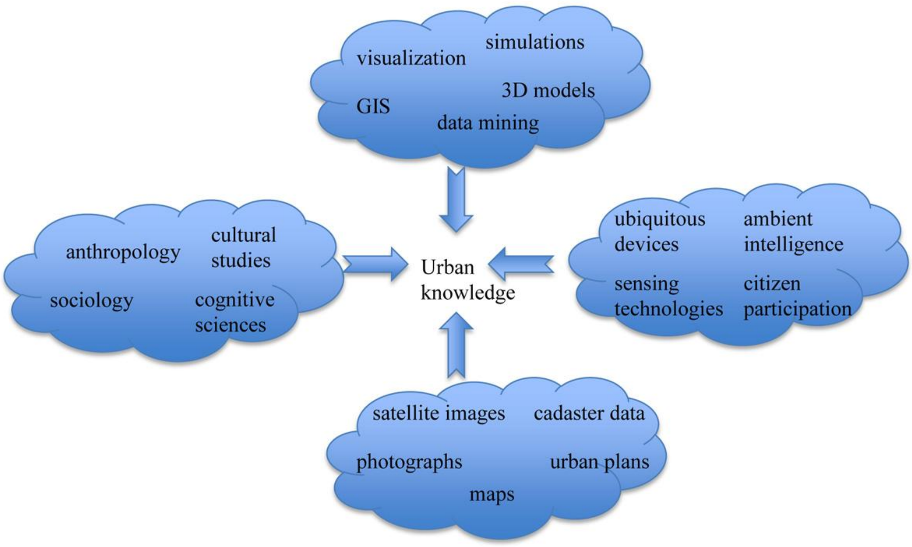

Fig. 1. Integrated urban knowledge

3 Predictive Modeling of Urban Dynamics

We can gather evidence and capture the footprints of urban change, though the explanation and interpretation used for modeling purposes must be interdisciplinary, theoretically and empirically plausible contributions. We have examined the ways in which urban changes might be influenced by various demographic, situational and environmental factors that characterize the context of interest. We argue that employment of intelligent technologies such as machine learning and data mining algorithms provide a potential solution to some of the challenges in urban modeling, especially automatic extraction and recognition of patterns in vast quantities of diverse types of evidence. Explanations of trends and manifestations derived from predictive model are more likely to match the reality, because the model accounts for deeper and richer relationships underneath data than simplistic statistical analysis. Validation is critical when modeling complex dynamic systems, hence the simulation tool have been recognized to serve both purposes, to capture the historical and empirical manifestations of urban change against which other approaches and models could be evaluated, and help in the interpretation and understanding of the studied behavior.

3.1 Application Overview: Editor Tools

A variety of image formats (e.g., orthographic image, scanned geographic map, cadaster data, AutoCAD image export) could be used as a spatial evidence of the urban state at a certain moment in time. Geo-referencing is a necessary requirement for their utilization. The image is overlaid with a cell grid with an adjustable cell size as shown in Figure 2. Cell type has a special importance in our simulations as a property that the model is trying to predict on the basis of past historical records. The timeline in the bottom ribbon represents the time period the predictive modeling spans across. Editor options give a user capability to set and assign the cells' properties and their values, while the Simulator tool is used for running simulations.

Color-coding is used for visual distinction between different cell types (Table 1); the assignment of colors is determined by the user. The process of cell description is a cumbersome and time-consuming process that needs to be repetitively performed for all available images of different time periods. Employment of AutoCAD image parser facilities the process of cell description (only those subject to change).

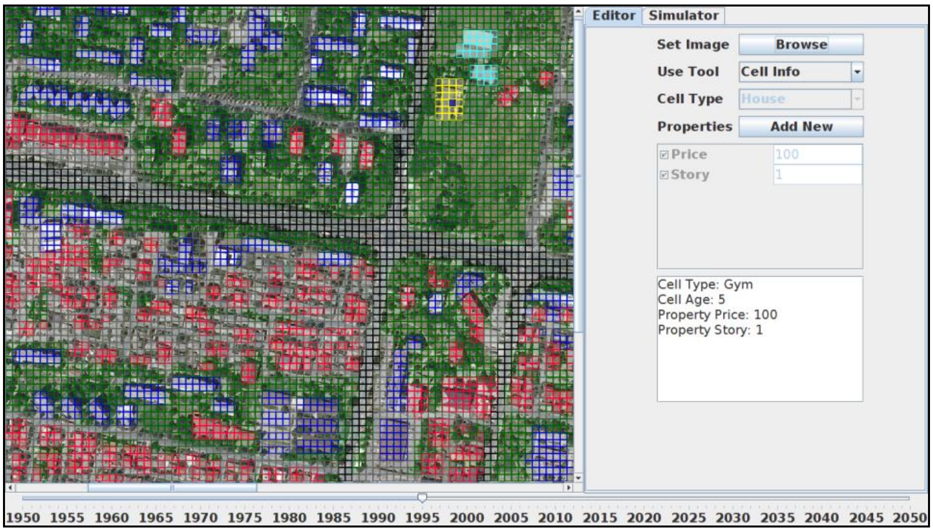

Fig. 2. Editor tool

support the representation of citizen-related data that have been shown to have an impact on the urban development (e.g., population size, age, income per family, preferences, proximity to work place).

Each cell is assigned with a variety of properties that should be proposed by domain experts and researchers from various fields. The set of selected categories hypothesized to have a prominent role in establishing appropriate urban knowledge could be broken down into three groups: (1) *urban factors*, (2) *socio-economic factors*, and (3) *population (individual and group) factors*.

Trends and circumstances associated with economic development, historic events, legislative initiatives, excessive urbanization, and local government policies are important since they are likely to affect the pace and direction of urban change. Socio-economic aspects may sometimes constrain, at other times may create new incentives, goals and directions for the urban development. The city's population size is included to support the representation of citizen-related data that have been shown to have an impact on the urban development. Color-coding is used for visual distinction between different cell types.

Table 1. Class legend.

	Class: Cell Type - Label
	Undeveloped - A
	Main Street - B
	Peripheral Street - C
	Industry - D
	House - E
	Education - F
	Park - G
	Residential + Commercial -H
	Government Offices - I
	Apartment Building - J
	Commercial - K
	Student Dormitory - L

3.2 Application Overview: Simulator Tool

We employ a variety of machine learning algorithms such as OneR [6], Ridor [7], PART [8], JRip [9], DecisionTable [10], J48 [11], SMO [12], IBk [13], KStar [14], NaiveBayes [15] and NNge [16]. The division of data into a training set, used to extract the patterns, and a test set, needed to evaluate the effectiveness of the predictions, could be adjusted for each run (Fig. 3). A set of features included in the predictive modeling includes not only the cell type and its properties, but also the types of its eight neighboring cells.

The performance results of the available algorithms are compared and the most suitable (independently or jointly used) set is selected. Summary results are presented in the bottom window, though the detailed results containing total number of instances, number of correctly and incorrectly classified instances, detailed accuracy metrics, confusion matrix and other statistics are available in the console output.

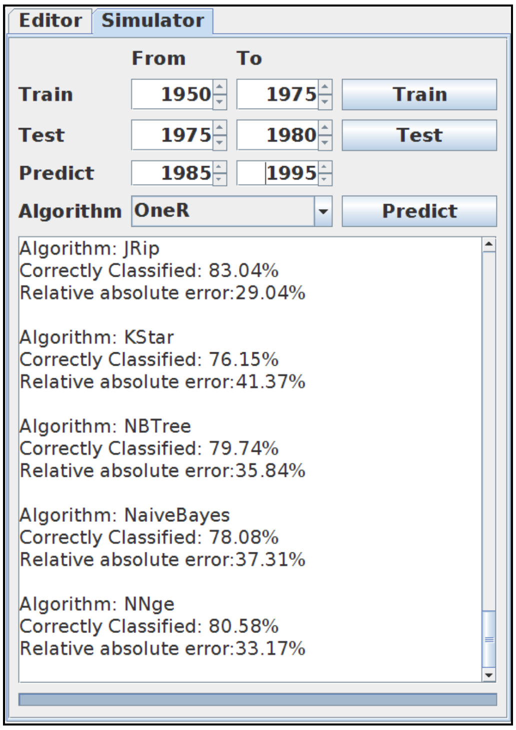

Fig. 3. Simulator tool

4 Case Study

The city of Skopje can be historically recognized as a traditional Balkan city that has been through a series of transformations. A succession of hallmark historic events and developments, followed by dissolution of preceding urban forms and patterns, has led to creation of complex urban strata that overlap and create the unique and complex imagery of the city. Different city fragments, each with a unique appearance, were in the focus of our study.

Our goal was to extract the rules of urban change, by modeling the states that different neighborhoods have undergone; from undeveloped land to dispersed residential houses to condensed areas with high-rise residential and commercial buildings. Some of the areas under investigation remain relatively compact throughout recent history; no extensive development until last decades. The study presented in this paper primarily focused on settlement in Skopje for the period from the 1960s. Orthographic images, scanned geographic maps, cadaster data and AutoCAD image exports were used (Fig. 4).

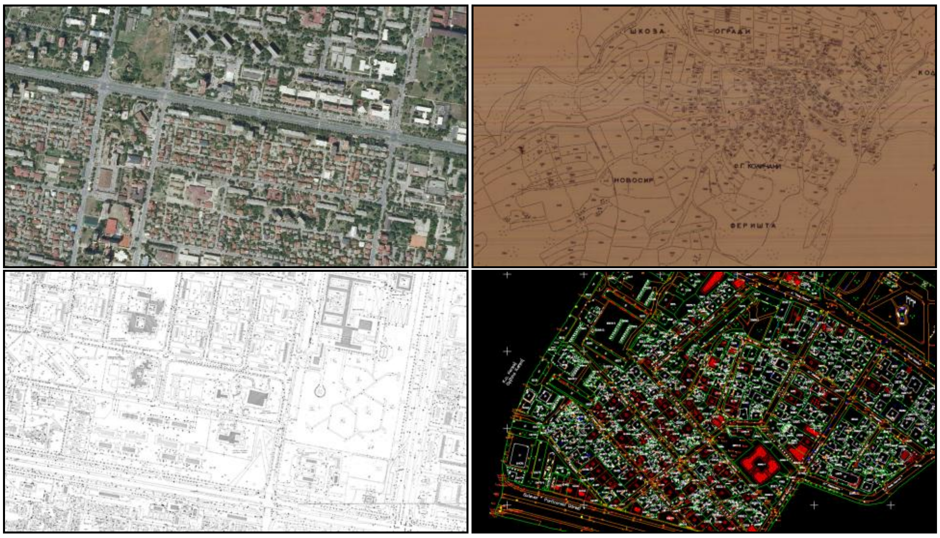

Fig. 4. Various urban data (1) Orthographic image (2) Historic geographic map (3) Cadastral record. (4) AutoCAD image export

To validate our models, we run a set of experiments to investigate how accurate the selected algorithms are at predicting patterns of urban change. The dataset was divided into a training sample of 1,110,786 cell instances and testing set containing 480,576 cells. The period from 1960 to 1996 was used to train the model, while the time period 1997 - 2013 was used to test the performance. We focus our discussion on the performance metrics obtained with PART algorithm, which has shown significant precision advantage. A total of 204 rules were generated. For illustration we have selected to show the rules regarding two types of residence dwellings, cell types House and Apartment Building. The rule-based analysis has revealed several patterns of housing residence sprawl shown in Fig. 5a. By urbanizing undeveloped land and taking over small parks, condensed and compact areas of houses emerged, reducing the space between a house and a peripheral street and diminishing the green zones. The emergent trends are in line with the rules derived with cellular automata model [17] although extended with new patterns, which could be clearly pinpoint to the exact time periods and related to socio-economic and population factors. While small isolated green zones were swallowed by housing development, and enlargement of existing commons (cell type Park) was detected as 4 rules (not shown).

The set of rules that pertain to transformation of cells into cell type Apartment Building shown in Fig. 5 are a clear evidence of the aggregation patterns resulting in clusters of buildings around cells already classified as buildings taking over neighboring houses or isolated park cells. A number of rules have provided a clear demonstration of transformations concerning main streets as well as the enlargement of industrial complexes (not shown in figures).

Fig. 5a. Transformation rules for cell type House

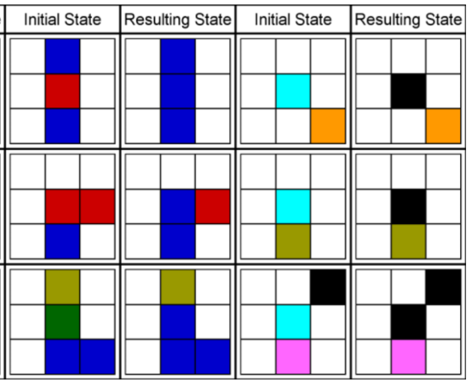

Fig. 5b. Transformation rules for cell type Apartment Building

The significance of the rules and their effectiveness in prediction of other time periods based on the rules discovered in historical data was also explored. Two types of analyses were performed differing in the number of test cells. The first one has concentrated only on the cells that undergone changes; in the second one all cells were considered, consequently the error numbers resulting from false predictions that a change will happen when in truth it did not occur were accounted for. Not surprisingly, most of the algorithms performed within the highest precision levels (98% - 99%) when all cells were considered due to the large difference between the total number of cells (in the order of 10^6) and the number of cells that experienced change (in the order of 10^3).

The confusion matrix for 8 out of 12 classification classes/cell type is presented in Table 2 and the corresponding performance metrics, namely accuracy and precision results for the selected classes are shown in Table 3. NBTree algorithm (a total of 660 leaves), J48 have also exhibited high levels of precision.

The use of data mining has allowed us to obtain effects and contributing factors, which are difficult to be derived by analytical resolution only. The changes revealed the significance of economic growth, demographics change, proximity, and neighborhood conditions on patterns of change in urban forms (type). Predictive modeling has

Table 2. Confusion matrix for 8 out of 12 classes.

Classified:	B	C	E	F	G	I	J	K
B	0	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	0
E	9	1	659	0	2	0	5	0
F	6	1	1	1011	0	0	7	0
G	0	0	0	0	385	0	0	0
I	0	0	0	0	0	60	0	0
J	0	1	2	1	0	0	958	0
K	0	0	0	0	0	2	0	886

Table 3. Precision and recall for each class; 1 and 2 denote the algorithm versions **with** and **without** automatic error corrections before each time steps, respectively.

	B	C	E	F	G	I	J	K
Precision 1	0	0	99.54	99.90	99.48	96.77	98.76	100
Recall 1	*	*	97.48	98.53	100	100	99.58	99.77
Precision 2	0	0	99.46	98.89	100	96.77	90.79	100
Recall 2	*	*	82.24	95.71	100	100	98.44	99.77

provided a starting point for multidisciplinary discussions, rules and trends that need to be interpreted by experts with a potential to be useful for decision making.

The performance metrics show great sensitivity to several features, the number of cells, cell size and time step in particular, which could be varied to achieve more efficient simulation cycles. Exploration into efficiency gains provisioned by new extensions to the set of semantic cell properties and/or cell agglomeration (depicting a region instead of a cell) invite further research efforts. Several aspects that would seem to be worthy investigating include the possible ways to account for incident local decisions and rare events, which seems to deteriorate the prediction performance for almost 20% and 3% for analysis number 1 and 2, respectively.

Our current efforts are directed toward predictive modeling of urban development by introducing variable sets of population and socio-economic variables - moving beyond the narrow focus on physical accessibility and the environment – and analyzing their correlation and significance within different machine learning algorithms.

5 Conclusions

We have discussed a set of basic questions that arise in the design and use of urban models, focusing on two main issues: the role of semantic enrichment and intelligent technologies in the process of discovery and explanation of emergent patterns of metropolitan change, and the predictive analysis as a way to make informative decision regarding urban planning. Our experience points at the relevance of applying intelligent technologies and predictive modeling that may help in understanding and representing urban phenomena. Still in the early phases of exploratory testing and model implementation, the proposed integrated data mining approach provide a promising start in building a foundation for urban modeling that aims at being both theoretically- and empirically-based.

Acknowledgements. This work was partially financed by the Faculty of Computer Science and Engineering at the "Ss. Cyril and Methodius" University.

References

1. Bretagnolle, A., Daude, E., Pumain. D.: From Theory To Modeling: Urban Systems As Complex Systems. CyberGe: European Journal of Geography 355, 1-17 (2005)
2. Hillier, B.: The Knowledge that Shapes the City: The Human City beneath the Social City. In Proceedings of the 4th Int. Space Syntax Symposium, pp. 1-20. London (2003)
3. Ettema, D., Jong, J., Timmermans, H., Bakema, A.:PUMA – Multi-agent modeling of urban systems. In Proceedings of CUPUM 05, Computers in Urban Planning and Urban Management, pp. 1-16. London (2005)
4. Rui, Y., Ban, Y.: Multi-agent simulation for modeling urban sprawl in the greater Toronto area. In Proceedings of 13th AGILE International Conference on Geographic Information Systems. Portugal, (2010)
5. Bonnefoy, J.L., Le Page, C., Rouchier, J., Bousquet F.: Modelling spatial practices and social representations of space using multi-agent systems. In: Ballot G., Weisbuch G. (eds.), Applications of Simulation to Social Sciences, pp. 115-168. Hermes Science, Paris (2000)
6. Holte, R.C.: Very simple classification rules perform well on most commonly used datasets. In: Machine Learning, vol. 11, pp. 63-91. (1993)
7. Gaines, B. R., Compton, P.: Induction of Ripple-Down Rules Applied to Modeling Large Databases. In: Journal of Intelligent Information Systems, 5(2), pp. 211–228. (1995)
8. Frank, E., Witten, I.H.: Generating Accurate Rule Sets Without Global Optimization. In: 15th International Conference on Machine Learning, pp. 144-151. (1998)
9. Cohen, W.: Fast effective rule induction. In Proceedings of the 12th International Conference on Machine learning, California (1995)
10. Kohavi, R.: The Power of Decision Tables. In: 8th European Conference on Machine Learning, pp. 174-189. (1995)
11. Quinlan, R.: C4.5: Programs for Machine Learning. Morgan Kaufmann, California (1993)
12. Platt J.: Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines. In: Schoelkopf, B., Burges, C., Smola, A. (eds.), Advances in Kernel Methods – Support Vector Learning. MIT Press (1998)
13. Aha, D., Kibler, D.: Instance-based learning algorithms. In: Machine Learning, vol.6, pp. 37-66. (1991)
14. Cleary, J.G., Trigg, L.E.:K* : An Instance- based Learner Using an Entropic Distance Measure. In Proceedings of the 12th International Conference on Machine learning, pp. 108-114. (1995)
15. John, G. H., Langley, P.: Estimating Continuous Distributions in Bayesian Classifiers. In Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence. pp. 338-345. Morgan Kaufmann, California (1995)
16. Brent M.: Instance-Based learning : Nearest Neighbor With Generalization. Master Thesis, University of Waikato, Hamilton, New Zealand (1995)
17. Marina, O.: Modeling Development of Urban Form in Residential Areas with System of Cellular automata. In Proceedings of 22nd Housing Research Conference ENHR 2010. Turkey, (2010)