# An overview of different data types and methods for urban land use analysis

Renato Andrade [1,2]
renatoandrade@dei.uc.pt

Ana Alves [1,2]
ana@dei.uc.pt

Carlos Bento [1]
bento@dei.uc.pt

[1] CISUC, Centre for Informatics and Systems, University of Coimbra, Polo II, 3030-290 Coimbra, Portugal
[2] Coimbra Institute of Engineering, Polytechnic Institute of Coimbra, 3030-199 Coimbra, Portugal

## Abstract

Modern planning and management of urban spaces is an essential topic for smart cities and depends on up-to-date and reliable information on urban land use. In the last years, driven by increased availability of geo-referenced data from social or embedded sensors and remote sensing (RS) images, various methods become popular for land use analysis. This paper addresses the various methods that are employed in this context, as well as data types needed for these techniques. From our study we concluded that even using the same methods and the same kind of datasets, results depend on spatial configuration of the data, accordingly to the specificity of each region. The work described in this paper is intended to provide relevant contributions to the selection of methods for knowledge discovery for city planning and management.

## 1 Introduction

With the recent and rapid development of cities, concerns with sustainability opened a new way for an essential field in recent studies: smart growth. It is an effort for better management of natural resources, by reducing and controlling its consumption [S+16]. The needs for urban land use planning and efficient management of urban areas have evidently become important [L+17]. These points are directly connected with the design and development of smart cities, converging to a common objective, which attempt to create a high quality of life for people in a more sustainable world. With attentions focused on urban spaces, land use analysis become essential.

Urban spaces have also gained focus because of issues related to urban expansion, hazard and pollution analysis, traffic control, well-being, population activity monitoring, construction projects, environmental preservation, economic analysis, as well as public health care and others topics. These subjects need essentially fine-grained maps to design and manage the work [L+17, Z+17b]. However, as urban areas change, maintaining maps and information about infrastructures and functional zones up-to-date is a challenge that research teams and public administration face daily, given the complexity of modern urban systems [Z+17b, Z+17a].

## 2 Data and methods for urban land use analysis

In this field, many methods can be applied based on different data types. An important task for researchers is to improve the results generated by these techniques. The integration of features extracted from various data types can to some extent show better results. We analysed a set of studies published in the last 5 years, identifying 16 different data types, as we can see in table 1, and 26 different methods as showed in table 2.

Table 1: Common data types used in scientific studies for urban land use analysis

| Data | Reference |
|---|---|
| Twitter activity | [FMFM14] |
| LBSN user activities | [ZUZ14, L$^+$17, GJC17] |
| Points of Interest (POI) | [ZUZ14, J$^+$15, Y$^+$15, Z$^+$17a, Z$^+$17b, L$^+$17, GJC17, Y$^+$17a] [XM18, S$^+$18, L$^+$18, Z$^+$19a] |
| Aggregate census employment data | [J$^+$15] |
| Boundaries of towns | [J$^+$15] |
| Remote images + extra attributes | [Dur15] |
| Taxi trajectories | [Y$^+$15, L$^+$18] |
| Pubic transit records | [Y$^+$15] |
| Remote sensing images | [Z$^+$17a, Z$^+$17b, L$^+$17, Y$^+$17b, S$^+$18, HZS18, L$^+$18, Z$^+$18b] [Z$^+$18a, D$^+$19, FZS19, Z$^+$19b] |
| Road network | [Z$^+$17b, S$^+$18] |
| LBSN users posts | [Z$^+$17b] |
| Traffic analysis zones (TAZ) | [Y$^+$17a] |
| Text messages | [XM18] |
| Building-level blocks | [XM18] |
| Road blocks | [HZS18, Z$^+$18b] |
| Origin-destination (OD) datasets | [Z$^+$19a] |

Table 2: Common methods used in scientific studies for urban land use analysis

| Method | Reference |
|---|---|
| Self-Organizing Maps (SOM) | [FMFM14] |
| Clustering | [FMFM14, ZUZ14, Dur15, GJC17, L$^+$17] [GJC17, Y$^+$17a, Z$^+$19a] |
| Nave Bayes | [ZUZ14] |
| Support Vector Machine (SVM) | [ZUZ14, Dur15, L$^+$17, D$^+$19] |
| Random Forest (RF) | [ZUZ14, Z$^+$17a, Z$^+$17b, Y$^+$17a, Y$^+$17b, XM18, Z$^+$19a] |
| Bayesian networks | [J$^+$15] |
| Tree-based learners | [J$^+$15] |
| Instance-based learners | [J$^+$15] |
| Rule-based learners | [J$^+$15] |
| Multiresolution segmentation | [Dur15, Z$^+$17a, S$^+$18, Z$^+$18b, D$^+$19] |
| Extreme learning machine (ELM) | [Dur15] |
| Latent Dirichlet Allocation (LDA) | [Y$^+$15, L$^+$17, GJC17, XM18] |
| Dirichlet Multinomial Regression (DMR) | [Y$^+$15] |
| Hierarchical semantic cognition (HSC) | [Z$^+$17a, Z$^+$18b] |
| Object-based classification | [Z$^+$17b, S$^+$18] |
| Probabilistic latent semantic analysis (pLSA) | [L$^+$17] |
| Word2Vec | [Y$^+$17a] |
| Google Inception v5 | [Y$^+$17b] |
| Semi-transfer DCNN | [HZS18] |
| Kernel density estimation | [L$^+$18] |
| Inverse hierarchical semantic cognition (IHSC) | [Z$^+$18a] |
| Object-based convolutional neural network | [Z$^+$18a] |
| Space-time fusion algorithm (ESTARFM) | [D$^+$19] |
| ResNet-50 DCNN | [FZS19] |
| Place2vec | [Z$^+$19a] |
| Joint Deep Learning (JDL) | [Z$^+$19b] |

# 3    Conclusions

In this paper, we discuss about knowledge discovery on urban land use and land cover, addressing the importance of functional regions in this context. Moreover, we analyzed several scientific studies related to this topic, making it possible to talk about the main challenges related to features selection. We also approached the main data types and the methods most frequently used in this specific field. During our analysis, we compared various works based on the types of data and the methods that were selected. We think this comparison is a source of new challenges, which we believe are essential to be considered in future work. In various cases, even using the same methods, for different regions, different authors arrived at different results and conclusions. Thus, we conclude that the results vary according to the method used, but also depend on the dataset and specificities of each region, due to factors such as construction patterns, population density and geography of the areas. Nevertheless, considering geographic data analysis as a specific topic of data analysis, it is important to remember that the results are directly related with data quality and granularity, but in this context, when using crowdsourced data for example, the spatial distribution of the data is also an essential factor to take into account.

Moreover, another consideration relates to the availability of data. During the study, we found the use of various data sources, and some of them are only available for some countries or regions. A very representative example of this situation is Weibo data, which is only available for China and building-level blocks, that is usually provided by public administration and is hardly available in various other locations. This limitation makes impossible or difficult to reproduce some studies in different locations.

In this research field, when talking about land use, a growing concern is related to the improvement of the accuracy of results, and therefore many authors have proposed the use of different data types together with remote sensing images. However, the use of innovative types of data, in many cases, did not result in a higher level of accuracy, compared to approaches that only use remote sensing images. This statement does not mean that combining data from multiple sources is not an important path to follow. From this observation we conclude that, depending on the chosen methodology, this wealth of data can improve the results obtained using remote sensing images or in cases where only one category of data is not enough to provide acceptable results.

# References

[D+19]      Z. Deng et al. Land use/land cover classification using time series Landsat 8 images in a heavily urbanized area. *Advances in Space Research*, 2019.

[Dur15]     S. S. Durduran. Automatic classification of high resolution land cover using a new data weighting procedure: The combination of k-means clustering algorithm and central tendency measures (KMC-CTM). *Applied Soft Computing Journal*, 35:136–150, 2015.

[FMFM14] V. Frias-Martinez and E. Frias-Martinez. Spectral clustering for sensing urban land use using Twitter activity. *Engineering Applications of Artificial Intelligence*, 35:237–245, 2014.

[FZS19]     E. Flores, M. Zortea, and J. Scharcanski. Dictionaries of deep features for land-use scene classification of very high spatial resolution images. *Pattern Recognition*, 89:32–44, 2019.

[GJC17]     S. Gao, K. Janowicz, and H. Couclelis. Extracting urban functional regions from points of interest and human activities on location-based social networks. *Transactions in GIS*, 21(3):446–467, 2017.

[HZS18]     B. Huang, B. Zhao, and Y. Song. Urban land-use mapping using a DCNN with high spatial resolution multispectral remote sensing imagery. *Remote Sensing of Environment*, 214(April):73–86, 2018.

[J+15]      S. Jiang et al. Mining point-of-interest data from social networks for urban land use classification and disaggregation. *Computers, Environment and Urban Systems*, 53:36–46, 2015.

[L+17]      X. Liu et al. Classifying urban land use by integrating remote sensing and social media data. *International Journal of Geographical Information Science*, 31(8):1675–1696, 2017.

[L+18]      X. Liu et al. Characterizing mixed-use buildings based on multi-source big data. *International Journal of Geographical Information Science*, 32(4):738–756, 2018.

[S+16]      R. Susanti et al. Smart Growth, Smart City and Density: In Search of The Appropriate Indicator for Residential Density in Indonesia. *Procedia - Social and Behavioral Sciences*, 2016.

[S+18]     J. Song et al. Mapping Urban Functional Zones by Integrating Very High Spatial Resolution Remote Sensing Imagery and Points of Interest. *Remote Sensing*, 10(11):1737, 2018.

[XM18]     H. Xing and Y. Meng. Integrating landscape metrics and socioeconomic features for urban functional region classification. *Computers, Environment and Urban Systems*, 72(February):134–145, 2018.

[Y+15]     N. J. Yuan et al. Discovering urban functional zones using latent activity trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 27(3):712–725, 2015.

[Y+17a]    Y. Yao et al. Sensing spatial distribution of urban land use by integrating points-of-interest and Google Word2Vec. *International Journal of Geographical Information Science*, 31(4):825–848, 2017.

[Y+17b]    Y. Yao et al. Sensing urban land-use patterns by integrating Google Tensorflow and scene-classification models. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(2W7):981–988, 2017.

[Z+17a]    X Zhang et al. Hierarchical semantic cognition for urban functional zones with VHR satellite images and POI data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 132:170–184, 2017.

[Z+17b]    Y. Zhang et al. The combined use of remote sensing and social sensing data in fine-grained urban land use mapping: A case study in Beijing, China. *Remote Sensing*, 9(9), 2017.

[Z+18a]    C. Zhang et al. An object-based convolutional neural network (OCNN) for urban land use classification. *Remote Sensing of Environment*, 216(June):57–70, 2018.

[Z+18b]    X. Zhang et al. Integrating bottom-up classification and top-down feedback for improving urban land-cover and functional-zone mapping. *Remote Sensing of Environment*, 212(Dec.):231–248, 2018.

[Z+19a]    W. Zhai et al. Beyond Word2vec: An approach for urban functional region extraction and identification. *Computers, Environment and Urban Systems*, 74(August 2018):1–12, 2019.

[Z+19b]    C. Zhang et al. Joint Deep Learning for land cover and land use classification. *Remote Sensing of Environment*, 221(November 2018):173–187, 2019.

[ZUZ14]    X. Zhan, S. V. Ukkusuri, and F. Zhu. Inferring Urban Land Use Using Large-Scale Social Media Check-in Data. *Networks and Spatial Economics*, 14(3-4):647–667, 2014.