Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2024) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

28th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2024)

Enhancing Land Use Patterns Understanding with Multi-Sensor, Multi-Temporal Metrics

Ricardo Muñoz-Cancino[a], Sebastián A. Ríos[a,*], Manuel Graña[b]

[a] *Business Intelligence Research Center (CEINE), Industrial Engineering Department, University of Chile, Beauchef 851, Santiago 8370456, Chile*
[b] *Computational Intelligence Group, University of Basque Country, 20018 San Sebastián, Spain.*

Abstract

Spatial-temporal land use patterns in urban environments are essential to understanding city dynamics. To uncover these patterns, many researchers have used the digital fingerprints of people's interaction with urban infrastructure, such as phone calls, facility check-ins, and geolocated social media activity. Despite multiple studies on the detection of land use patterns in urban environments, the need for more consensus on the appropriate metrics to define which set of patterns best describes the dynamics of a city remains a significant limitation. This evaluation is often subjective and depends on the researcher's in-depth knowledge of the study area, which makes an extensive comparison of multiple cities difficult.

This paper introduces a novel set of metrics to determine the patterns that best represent urban activity, diminishing subjective interpretations. Our methodology, which tests our metrics on land use patterns obtained from a dynamic topic model, is a fresh approach to the field. To apply our methodology, we use a dataset of human urban activities collected over 17 years in cities with more than 1 million inhabitants or country capitals. Our results demonstrate that these metrics are a starting point for understanding, analyzing, and choosing the land use patterns that best describe the dynamics and use given to urban space. The practical implications of our research are significant, as it can guide decision-making processes and contribute to the sustainable development of urban areas. However, it is important to highlight that there is still work to be done to reach a consensus on the optimal metrics to evaluate these patterns.

Keywords: Land Use Patterns; Land Use Assessment; Multi-Sensor; Multi-Temporal

1. Introduction

The convergence of digital traces and geo-crowdsourcing data offers a unique opportunity to understand human behavior in urban environments and planning contexts [4, 27, 29, 31]. These datasets can provide useful insights

* Corresponding author.
E-mail address: srios@dii.uchile.cl

for effective transportation management [12, 23, 28], for addressing natural disaster management and mitigation [5, 6, 30, 33], analyze patterns of terrorist activities [2, 10, 26], among others. These applications are diverse and have significant implications for society. In parallel, we have witnessed a notable expansion of technology in recent decades, with advances that have led to its ubiquity in our daily lives [1]. It is common for people to have multiple technological devices that function as sensors, capturing various aspects of their daily activities. Mobile phones, ubiquitous in today's society, record communication patterns, track location through GPS [34], and facilitate access to social networks where ideas and experiences are shared. Given this context, taking advantage of this wealth of data is essential to strengthen urban planning and the management of complex infrastructures. However, accurately assessing the land use patterns that emerge from the analysis of these data remains a challenge, with subjectivity and lack of consensus on appropriate metrics as important limitations. A widely used evaluation methodology of land use patterns relies on the expertise and deep knowledge of the researcher. This approach allows for a more subjective evaluation of the land uses and their relevance to the study's specific context. However, it can be challenging for researchers to obtain this knowledge when the study subject areas are extensive or unfamiliar. In such cases, alternative evaluation methods may need to be considered, such as consulting with local experts or using additional data sources to understand the area better. The most popular metric used is the cosine distance, which is used as a dissimilarity measure and is most commonly used to evaluate the similarity between topics [7, 14, 20, 32]. Therefore, in this study, we aim to address the following research question: *What objective metrics can be defined, calculated using multi-sensor and multi-temporal data, and applied to select the best representation of spatio-temporal land use patterns?*

We utilize the Geotagged Digital Traces dataset [19] used in our previous research to address these research questions effectively. This dataset includes 32 million records from various fingerprint datasets covering mixed urban activities. This collection includes activities such as geolocated tweets, images, and check-ins on various social media platforms. The information collected spans more than 17 years, allowing us to model temporal dependency and explore the dynamic evolution of activity patterns.

The paper is structured as follows: Section 2 provides background details on previous work. Section 3 details the proposed metrics, methodology, and applied computational methods. Section 4.1 describes the dataset used in this study. Section 4 provides details of the experimental setup. Section 5 presents and discusses the results. Finally, Section 6 offers our conclusions derived from the findings and suggests ideas for future research efforts.

2. Background and Related Work

The validation of land use patterns in urban environments has been an area of active research, especially with the exponential growth of data from various technological and social sources. These data provide a detailed view of human activities and land use behaviors in urban contexts, offering opportunities to improve urban planning and infrastructure management. In this sense, various approaches and techniques have been developed to validate and understand land use patterns. For example, some researchers have used aggregation and dispersion models to analyze movement histories and regional characteristics of human mobility [9]. These approaches offer a detailed understanding of the historical trajectories of human mobility and the distinctive characteristics of different urban regions. Others have employed spectral clustering techniques to model patterns of crowd activity in multiple cities [22], allowing the identification of large-scale spatial patterns of human activity. Additionally, methods that leverage data from social media and mobile applications have been explored to validate and understand land use patterns. For example, techniques such as the mean-shift algorithm have been used to identify landmarks using geolocated photo data [3]. Similarly, using geolocated tweets as a complementary data source for urban planning applications has been investigated, using techniques such as self-organizing maps, Voronoi tessellations, and K-means [8]. In urban planning and infrastructure management, validating land use patterns is essential to understand human interactions with public infrastructure and predict future infrastructure needs [27, 31]. However, the need for more objective and precise data has traditionally limited this process. In response, modern approaches have been developed using human activity records from telecommunication companies and satellite image data to extract aggregated information on land use in urban areas [11, 16, 24, 29]. As we have shown, there are several approaches to characterizing a city based on urban activities, resulting in multiple ways of evaluating the quality of the resulting patterns. A commonly used approach is to evaluate the consistency of the identified patterns or clusters using metrics such as cosine similarity [7, 14, 20, 32] or information entropy [15]. When the discovered activity patterns include a spatial component, geospatial metrics, such as pattern distribution and coverage over the study area, are used to evaluate the results [13, 25].

3. Proposed metrics for assessing Multi-Sensor and Multi-Temporal Land Use Patterns

There are various approaches to characterizing land use patterns based on urban activities. As a result, there are multiple ways to evaluate the quality of the results of a land use detection methodology. However, existing shared foundations allow us to establish specific guidelines on the quality of the land use patterns discovered. One commonly used methodology is to assess the consistency of identified patterns or clusters using metrics [7]. When the discovered land use patterns present a spatial component, geospatial metrics, such as pattern distribution and coverage over the study area, are used to evaluate the results. Our research has led us to identify desired properties for temporal land use patterns. When combined with the corresponding metrics, these properties provide a practical and standardized framework for assessing the quality of the patterns discovered. This framework is a tool that can be readily applied in urban planning and land use, enabling professionals to make informed decisions based on reliable data. To mathematically represent land-use patterns, we will use the exact definitions that were used in our previous research [17, 18, 21, 24]. Assuming a similarity function $sim(x, y)$ between two land use patterns, the desired properties of temporal land use patterns are the following:

- **Intratemporal Similarity:** One of the primary expected outcomes is that the patterns explain different activities carried out in a urban environment. Therefore, land use patterns will be as dissimilar as possible between them. Formally speaking, given a set of temporal land use patterns $\mathcal{K}$, a time-slice partition $\mathcal{S}$ and K temporal land use patterns, the Intratemporal Similarity takes the form:

$$IntraSim(\mathcal{K}) = \frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \frac{2}{K(K-1)} \sum_{k=0}^{K-1} \sum_{l=0}^{K-1} \mathbb{1}_{k \neq l} \cdot sim(AT_{s,k}, AT_{s,l}) \tag{1}$$

  Where $\mathbb{1}_{k \neq l}$ is the indicator function that values 1 when the predicate is true, and zero elsewhere. The optimal set of land use patterns will minimize the Intratemporal Similarity. When the function used to compare land use patterns is a distance function, the minimization problem becomes a maximization problem.

- **Intertemporal Stability:** When an urban environment is analyzed over time, the measured urban activity behavior is expected to stay the same over time. We will measure these gradual land-use behavior differences from the changes between the same land-use pattern in adjacent time slices. Formally speaking, given a set of temporal land use patterns $\mathcal{K}$, a time-slice partition $\mathcal{S}$, and K temporal land use patterns, the Intertemporal Stability takes the form:

$$InterSta(\mathcal{K}) = \frac{1}{K} \sum_{k=0}^{K-1} \frac{1}{2|\mathcal{S}|-2} \sum_{s=0}^{|\mathcal{S}|-1} \sum_{u=0}^{|\mathcal{S}|-1} \mathbb{1}_{|s-u|=1} \cdot sim(AT_{s,k}, AT_{u,k}) \tag{2}$$

  The optimal set of land use patterns will maximize the Intertemporal Stability. The maximization problem becomes a minimization problem when the function used to compare land use patterns is a distance function.

- **Topic Consistency:** Urban activity is based on their routine inhabitants, which is why there is a certain regularity in the activities carried out on working days and during the weekend. This consistency can be observed in the empirical results obtained in multiple studies. Taking this into account, we propose the following metric of topic coherence of weekly land use patterns $AT_{s,k}$:

$$TC(AT_{s,k}) = \frac{5}{7} \sum_{i,j \in weekdays} \mathbb{1}_{i \neq j} \cdot sim(AT_{s,k,i}, AT_{s,k,j}) + \frac{2}{7} \sum_{i,j \in weekend} \mathbb{1}_{i \neq j} \cdot sim(AT_{s,k,i}, AT_{s,k,j}) \tag{3}$$

Where $AT_{s,k,i}$ correspond to components of day i in the activity topic $AT_{s,k}$, i.e, the first 24 components of $AT_{s,k}$ represents a Monday ($i = 1$), the next 24 components a Tuesday ($i = 2$), and so on.

- **Topic Smoothness:** Urban activity is carried out continuously, without significant restrictions that suddenly limit all activity. For this reason, land-use patterns with smooth changes are preferred to those with high volatility during the day. We propose a simple and effective method to measure the smoothness of an activity topic.

$$TS(AT_{s,k}) = \sqrt{\frac{1}{T-2} \sum_{i=1}^{T-1} (d_{s,k,i} - \bar{d}_{s,k})^2} \tag{4}$$

Where $AT_{s,k} \in \mathcal{R}^T$, the difference vector $d_{s,k,i} = AT_{s,k,i+1} - AT_{s,k,i},\ i = 1, \cdots, T-1$ and $\bar{d}_{s,k} = \frac{1}{T-1} \sum_{i=1}^{T-1} d_{s,k,i}$. Optimal land use patterns will minimize this metric.

These metrics provide an objective framework for comparing the results of different land-use modeling processes, ensuring that our evaluation method is comprehensive and rigorous. This allows us to accurately compare the temporal city activity patterns discovered by diverse computational approaches. Unless stated otherwise, in this paper, the above metrics were calculated using the cosine distance as the dissimilarity measure, commonly used to evaluate the similarity between patterns [7, 14, 20, 32]. If two patternsare identical, their cosine distance equals zero, reflecting their high similarity.

4. Experimental Setup

This study aims to apply the proposed metrics to assess the quality of land use patterns obtained from various cities worldwide over a 17-year study period detailed in 4.1. This methodology utilized land-use patterns obtained in previous research [17] using Dynamic Topic Modeling (DTM) and other algorithms. Therefore, we will follow the following steps to determine the best representation of land use.

4.1. Data Description

The Geotagged Digital Traces dataset [19] is used in this study. It comprises approximately 32 million geotagged urban activities obtained from social platforms such as Twitter, Foursquare, Yelp, Flickr, Gowalla, Brightkite, and Weeplaces, collected over 17 years. Each urban activity is characterized by its geographical location and a time stamp. This dataset is available to ensure the study's reproducibility. In this work, we use the same definitions and assumptions used in our previous research [17, 18, 21], where each urban activity was assigned to the nearest city, as long as it is less than 30 km from the city center.

4.2. Experimental setup for assessing city land-use patterns

- The initial step compares outcomes from different static models, such as LDA, K-Means, k-Shape, and Time Series K-Means, to induce land-use patterns. Each time slice requires a separate model due to the lack of temporal consideration.
- The unsupervised static modeling assigns varied labels to similar patterns across time slices, possibly representing the evolution of the same land use. To mitigate this, a heuristic algorithm is employed to assign identical labels to the most similar land use patterns from different time slices, with the findings detailed in section 5.1.
- The dataset is divided into one-year and three-year time slices to establish the most effective temporal grouping. Results of this comparison are detailed in section 5.2. Initially, models are trained using one-year intervals, generating 17 pattern subsets, one for each year. Afterward, models are trained using three-year intervals. In both cases, the number of land-use patterns ranges from $k = 2$ to $k = 5$.
- Finally, the model that most accurately characterizes land use is chosen, and the optimal number of land use patterns is identified. The findings of this comparison and analysis are outlined in section 5.3.

5. Results and Discussion

This section presents results from applying the proposed metrics to a geo-tagged urban activities dataset to identify Land Use Patterns. It includes analyses on temporal matching (See 5.1), temporal aggregation (See 5.2), optimal model selection, and determining the number of activity patterns (See 5.3). Section 5.4 discusses the final model and its real-life implications.

5.1. Assessing the temporal matching heuristic

This subsection evaluates the effectiveness of a temporal matching heuristic in aligning land use patterns across sequential time slices. Traditional topic modeling processes are unsupervised, so similar topics across time slices may receive different labels. A greedy heuristic is applied to match similar topics between consecutive time slices and assign them the same label to address this.

Table 1 presents the results, showing the percentage increase in Intertemporal Stability achieved through topic relabeling using the heuristic compared to unsupervised labeling. The table indicates the resolution of time slice aggregation and the topic modeling process used. The results demonstrate significant improvement in Intertemporal Stability for most processes with the heuristic, confirming its effectiveness in aligning activity patterns across time. Hence, the relabeled activity topic labels will be utilized in subsequent analyses.

Table 1. The result of temporal matching heuristic of land use patterns. Time matching impact is measured using the Intertemporal Stability with cosine similarity. Time Aggregation corresponds to the two temporal aggregation scenarios used. Model column shows the algorithm used to obtain the patterns. Topics indicate the number of patterns obtained with each algorithm, varying between $K = 2$ and $K = 5$. Each value in the cells is the percentage increment between the heuristic relabeling and the original topic labels

Time Aggregation	Model	#Topics			
		2	3	4	5
one-year	K-Means	0.97%	7.57%	5.76%	16.46%
	k-Shape	1.09%	2.45%	0.78%	1.70%
	LDA	0.00%	0.00%	0.29%	0.09%
	TS K-Means	0.98%	0.87%	2.62%	2.41%
three-year	K-Means	3.22%	4.33%	3.53%	4.21%
	k-Shape	1.62%	0.05%	0.48%	0.89%
	LDA	0.00%	0.00%	0.35%	0.00%
	TS K-Means	3.67%	2.70%	4.99%	6.52%

5.2. Optimal Time-slices Aggregation

This section compares two time-slice resolution aggregations for training land-use pattern detection models. Table 2 compares the two selected time resolutions, showcasing outcomes for proposed and traditional models. Metrics such as Intertemporal Stability, Intratemporal Similarity, Topic Smoothness, and Topic Consistency are computed for each set of activity patterns obtained. For each metric, three corresponding columns are presented: average over three-year time slices, average over one-year time slices, and the percentage variation between both, calculated based on the three-year results.

Results show that Intertemporal Stability is generally lower for three-year time-slices, indicating greater stability over time than shorter time-slices. However, the Intratemporal Similarity index for three-year time-slices is lower, suggesting less diversity within topics compared to one-year time-slices. Nevertheless, this result is influenced by high volatility and inconsistency in obtained topics, with Topic Smoothness notably lower for three-year time-slices, while Topic Consistency is higher.

Given the findings, we will continue the analysis using three-year time-slice datasets. This choice is based on the better Intertemporal Stability, Topic Smoothness, and Topic Consistency demonstrated by three-year time slices. This approach will enable us to comprehensively and precisely understand land uses and their evolution over time, further advancing our research.

Table 2. Time-slices aggregation comparison. Inter-temporal Stability, Intra-temporal Similarity, Topic Smoothness, and Topic Consistency validation metrics for one-year (1Y) and three years (3Y) time-slice dataset partitions. Topics indicate the number of patterns obtained with each algorithm, varying between K=2 and K=5. Model column shows the algorithm used to obtain the patterns. DIFF corresponds to the percentage variation between both columns (3Y and 1Y) based on the three-year results.

Topics	Model	Intertemporal Stability			Intratemporal Similarity			Topic Smoothness			Topic Consistency		
		3Y	1Y	DIFF	3Y	1Y	DIFF	3Y	1Y	DIFF	3Y	1Y	DIFF
$k = 2$	K-Means	0.31	0.31	-0.58%	0.36	0.42	-13.6%	5.25	15.39	-65.8%	0.82	0.75	8.6%
	k-Shape	0.08	0.14	-40.3%	0.06	0.15	-58.5%	1.78	3.44	-48.1%	0.97	0.92	5.0%
	TS K-Means	0.31	0.29	6.5%	0.36	0.39	-8.4%	5.25	11.15	-52.8%	0.82	0.76	7.1%
	LDA	0.07	0.07	-2.42%	0.03	0.08	-60.3%	1.93	2.92	-33.9%	0.96	0.92	4.9%
	DTM	0.00	0.07	-89.3%	0.32	0.43	-24.4%	1.78	3.78	-52.9%	0.98	0.88	11.0%
$k = 3$	K-Means	0.29	0.41	-27.4%	0.31	0.48	-35.1%	5.37	18.66	-71.1%	0.81	0.65	24.1%
	k-Shape	0.12	0.14	-15.5%	0.05	0.16	-68.5%	1.82	3.62	-49.7%	0.97	0.89	9.0%
	TS K-Means	0.29	0.38	-21.8%	0.31	0.46	-32.0%	5.37	14.04	-61.6%	0.81	0.67	21.0%
	LDA	0.05	0.12	-60.2%	0.12	0.23	-49.0%	2.51	4.68	-46.2%	0.92	0.85	7.8%
	DTM	0.00	0.09	-93.4%	0.40	0.49	-17.4%	2.00	5.91	-66.0%	0.97	0.77	25.2%
$k = 4$	K-Means	0.34	0.43	-20.5%	0.35	0.47	-24.9%	6.47	20.5	-68.4%	0.76	0.65	17.1%
	k-Shape	0.19	0.18	6.3%	0.12	0.15	-21.8%	2.88	3.6	-20.0%	0.91	0.88	3.5%
	TS K-Means	0.34	0.41	-15.5%	0.35	0.46	-22.8%	6.47	15.95	-59.4%	0.76	0.66	15.4%
	LDA	0.04	0.14	-71.8%	0.15	0.26	-42.3%	2.95	5.17	-42.9%	0.89	0.83	7.7%
	DTM	0.00	0.1	-94.7%	0.42	0.52	-18.1%	2.24	6.72	-66.6%	0.96	0.74	30.6%
$k = 5$	K-Means	0.35	0.57	-37.4%	0.37	0.61	-38.8%	6.38	25.03	-74.4%	0.73	0.54	35.5%
	k-Shape	0.15	0.24	-34.8%	0.10	0.21	-49.5%	2.68	5.33	-49.7%	0.92	0.85	8.5%
	TS K-Means	0.31	0.46	-32.6%	0.31	0.46	-32.5%	5.64	17.47	-67.6%	0.78	0.65	19.1%
	LDA	0.02	0.14	-84.3%	0.14	0.26	-45.9%	2.88	5.47	-47.2%	0.89	0.83	7.3%
	DTM	0.00	0.09	-95.1%	0.37	0.63	-39.8%	2.58	9.22	-72.0%	0.95	0.71	33.2%

5.3. Selection of optimal model and number of land use patterns

This section evaluates models using proposed metrics to identify the optimal land-use representation. Using the same measures, the optimal number of topics is also determined. Models are trained using three-year time slices and the temporal matching heuristic.

The Intertemporal Stability metric is considered to determine the optimal model. Table 3 presents results for models trained with three-year time slices. Each row represents a trained algorithm, and each column shows index values as the number of topics varies from $k = 2$ to $k = 5$. Regularity in index values is observed within each algorithm, while considerable differences exist between algorithms. K-Means-based models exhibit more unstable topics over time, whereas LDA and DTM models demonstrate stability. Across all scenarios, the methodological proposal using DTM achieves the best results regarding the Intertemporal Stability metric.

Secondly, we consider the Intratemporal Similarity, finding that the differences between DTM and the rest of the models are less notorious than in the previous case. This data can be seen in Table 4, where DTM obtains similar results to those obtained by K-Means and TS K-Means. However, DTM obtains a better Intratemporal Similarity Index except for $k = 2$ topics extracted. The best results for DTM are obtained by extracting $k = 3$ and $k = 4$ Topics.

Table 3. Intertemporal Stability using cosine distance. Model column shows the algorithm used to obtain the patterns. Topics indicate the number of patterns obtained with each algorithm, varying between K=2 and K=5. Each value in the cells corresponds to the Intertemporal Stability.

Model	#Topics			
	2	3	4	5
K-Means	0.31	0.29	0.34	0.35
k-Shape	0.08	0.12	0.19	0.15
TS K-Means	0.31	0.29	0.34	0.31
LDA	0.07	0.05	0.04	0.02
DTM	0.00	0.00	0.00	0.00

Table 4. Intratemporal Similarity using cosine distance. Model column shows the algorithm used to obtain the patterns. Topics indicate the number of patterns obtained with each algorithm, varying between K=2 and K=5. Each value in the cells corresponds to the Intratemporal Similarity.

Model	#Topics			
	2	3	4	5
K-Means	0.36	0.31	0.35	0.37
k-Shape	0.06	0.05	0.12	0.10
TS K-Means	0.36	0.31	0.35	0.31
LDA	0.03	0.12	0.15	0.14
DTM	0.32	0.40	0.42	0.37

Thirdly, we consider Topic Consistency results presented in Table 5. It shows that the DTM model has the highest Topic Consistency among all models tested. It is important to note that Topic Consistency measures the regularity of patterns obtained by comparing days of the week so that the components of an urban activity topic should stay the same from one day to the next. This expected behavior is because the activities carried out in the city reflect the routine of the people who inhabit them. Unlike previous indices, this metric is calculated using cosine similarity instead of cosine distance. In this way, the more consistent the topics obtained, the indicator will be closer to one. However, it is worth mentioning that as the number of activity patterns increases, topic consistency tends to decrease, but DTM exhibits a minor variation.

Table 5. Topic Consistency using cosine similarity. Model column shows the algorithm used to obtain the patterns. Topics indicate the number of patterns obtained with each algorithm, varying between K=2 and K=5. Each value in the cells corresponds to the Topic Consistency.

Model	#Topics			
	2	3	4	5
K-Means	0.82	0.81	0.76	0.73
k-Shape	0.97	0.97	0.91	0.92
TS K-Means	0.82	0.81	0.76	0.78
LDA	0.96	0.92	0.89	0.89
DTM	0.98	0.97	0.96	0.95

Finally, the analysis of Topic Smoothness, as displayed in Table 6, indicates that K-Means and Time-Series K-Means reach the highest values for this indicator among all models analyzed. This result indicates that patterns obtained from these methods exhibit significant variations between consecutive hours. This behavior is not expected to be observed in an Urban Activity Pattern where gradual changes are expected rather than drastic fluctuations from one hour to the next. Additionally, the remaining models produce results of similar magnitude, with DTM consistently outperforming others in nearly all scenarios. It should be noted that as the number of urban patterns increases, the Topic Smoothness value tends to rise.

Regarding the optimal number of topics, from the results obtained in this computational exploration, we have determined that $k = 3$ is the most appropriate number of urban land use patterns representing the behavior of the cities included in this study. This decision is based on the Intertemporal Stability and Intratemporal Similarity measurements. Also, it considers the trade-off of increasing the number of urban activities that were observed when analyzing Topic Consistency and Topic Smoothness.

Table 6. Topic Smoothness. Model column shows the algorithm used to obtain the patterns. Topics indicate the number of patterns obtained with each algorithm, varying between K=2 and K=5. Each value in the cells corresponds to the Topic Smoothness.

Model	#Topics			
	2	3	4	5
K-Means	5.25	5.37	6.47	6.38
k-Shape	1.78	1.82	2.88	2.68
TS K-Means	5.25	5.37	6.47	5.64
LDA	1.93	2.51	2.95	2.88
DTM	1.78	2.00	2.24	2.58

5.4. Multi-sensor and multi-temporal land-use patterns in the optimal model

This section presents land-use patterns obtained by the optimal model, depicted in Figure 1. The figure comprises three columns, each representing one of the extracted topics (Land Uses), and six rows, each corresponding to a subset of three topics obtained for a particular time slice identified by the time range shown in the column corresponding to Topic 0. The interpretation of land-use patterns is provided as follows:

- **Topic 0:** This pattern is characterized by regular workday behavior, with activity increasing throughout the day and peaking twice, at noon and 9 pm. This pattern, also observed in previous studies, indicates Leisure and commerce activities. On weekends, both peaks are preserved, with the noon peak being less pronounced, giving way to increased nighttime activity, including an intense Sunday nightlife.
- **Topic 1:** There is relatively low activity during workdays, with a notable increase on weekends. During workdays, activity gradually increases from 9 am to 9 pm without significant variations. Activity rises from 9 am on weekends, peaking at 3 pm before sharply declining, especially on Sundays.
- **Topic 2:** Displays distinct behaviors during workdays and weekends. On workdays, activity is concentrated between 9 am and 6 pm, with a slight dip around noon. This pattern resembles office-area activity [20, 24]. On weekends, activity peaks in the morning but declines steadily throughout the day.

These interpretations offer insights into the temporal and behavioral dynamics of urban land-use patterns captured by the optimal model

6. Conclusions and Future Work

This paper addresses important questions in urban planning, traffic management, and policy design. Our proposed methodology is a significant step toward closing the gap in assessing land use patterns. By introducing four metrics, we aim to reduce dependence on expert knowledge, thereby enhancing objectivity and trustworthiness in pattern selection. In this paper, we have defined and calculated objective metrics using multi-sensor and multi-temporal data. We have demonstrated that these metrics can be effectively applied to select a representation of spatial and temporal land use patterns. Our findings demonstrate that the proposed metrics effectively facilitate the identification of a representative set of land use patterns. This approach reduces the dependence on expert manual analysis of each set of patterns, thus minimizing the dependence on domain-specific knowledge. Metrics allow for more systematic and quantifiable evaluation, ensuring the selection process is based on insights derived from data rather than subjective interpretations.

Through strict analysis, we have established that these metrics offer a methodical approach to evaluating and selecting patterns representing urban activity dynamics. This methodology reduces dependence on expert knowledge, which has traditionally been fundamental in analyzing land use patterns. Furthermore, our study demonstrates the relevance of these metrics in various urban environments and time scales. Whether analyzing multi-sensor data or considering multi-temporal patterns over several years, our metrics consistently deliver reliable results.

However, it is important to highlight that there is still work to reach a consensus on the optimal metrics to evaluate these patterns. Although our proposed metrics have shown promising results, more research is needed to refine and validate them in different contexts and datasets. Future work in this area could focus on conducting comparative studies involving a broader range of evaluation metrics and datasets.

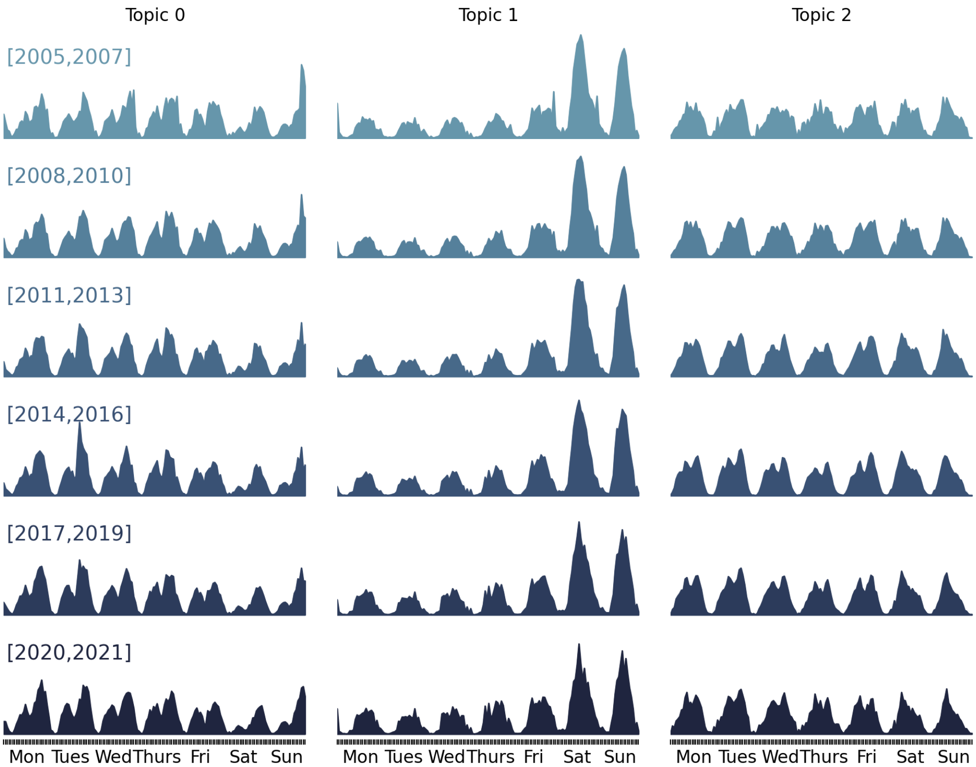

Fig. 1. Multi-sensor and multi-temporal land-use patterns obtained using Dynamic Topic Models in [17]. Rows correspond to three year time slices, columns correspond to the topic profiles. Three topic model was deemed optimal in our objective assessment.

References

[1] BankMyCell, 2022. How many smartphones are in the world? Retrieved from https://www.bankmycell.com/blog/how-many-phones-are-in-the-world. Accessed October 13, 2022.

[2] Berube, M., Tang, T.U., Fortin, F., Ozalp, S., Williams, M.L., Burnap, P., 2020. Social media forensics applied to assessment of post–critical incident social reaction: The case of the 2017 manchester arena terrorist attack. Forensic Science International 313, 110364. doi:10.1016/j.forsciint.2020.110364.

[3] Crandall, D.J., Backstrom, L., Huttenlocher, D., Kleinberg, J., 2009. Mapping the world's photos, in: Proceedings of the 18th international conference on World wide web, WWW 2009, Madrid, Spain, April 20-24, 2009, ACM, New York, NY, USA. pp. 761–770. doi:10.1145/1526709.1526812.

[4] Dembski, F., Wössner, U., Letzgus, M., Ruddat, M., Yamu, C., 2020. Urban digital twins for smart cities and citizens: The case study of herrenberg, germany. Sustainability 12. doi:10.3390/su12062307.

[5] Fan, C., Zhang, C., Yahja, A., Mostafavi, A., 2021. Disaster city digital twin: A vision for integrating artificial and human intelligence for disaster management. International Journal of Information Management 56, 102049. doi:10.1016/j.ijinfomgt.2019.102049.

[6] Farahmand, H., Wang, W., Mostafavi, A., Maron, M., 2022. Anomalous human activity fluctuations from digital trace data signal flood inundation status. Environment and Planning B: Urban Analytics and City Science 49, 1893–1911.

[7] Frias-Martinez, V., Frias-Martinez, E., 2014. Spectral clustering for sensing urban land use using twitter activity. International Scientific Journal Engineering Applications of Artificial Intelligence 35.

[8] Frias-Martinez, V., Soto, V., Hohwald, H., Frias-Martinez, E., 2013. Sensing urban land use with twitter activity. Telefonica Research, Madrid, Spain .

[9] Fujisaka, T., Lee, R., Sumiya, K., 2010. Exploring urban characteristics using movement history of mass mobile microbloggers, in: Proceedings of the Eleventh Workshop on Mobile Computing Systems & Applications, HotMobile '10, Annapolis, Maryland, USA, February 22-23, 2010, ACM, New York, NY, USA. pp. 13–18. doi:10.1145/1734583.1734588.

[10] Garcia, D., Rime, B., 2019. Collective emotions and social resilience in the digital traces after a terrorist attack. Psychological Science , 617–628doi:10.1177/0956797619831964. pMID: 30865565.

[11] Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.L., 2008. Understanding individual human mobility patterns. Nature 453, 779–782. doi:10.1038/nature06958.

[12] Graff, M., Moctezuma, D., Miranda-Jimenez, S., Tellez, E.S., 2022. A python library for exploratory data analysis on twitter data based on tokens and aggregated origin–destination information. Computers & Geosciences 159, 105012. doi:10.1016/j.cageo.2021.105012.

[13] Herold, M., Couclelis, H., Clarke, K.C., 2005. The role of spatial metrics in the analysis and modeling of urban land use change. Computers, Environment and Urban Systems 29, 369–399. doi:10.1016/j.compenvurbsys.2003.12.001.

[14] Hu, N., Legara, E.F., Lee, K.K., Hung, G.G., Monterola, C., 2016. Impacts of land use and amenities on public transport use, urban planning and design. Land Use Policy 57, 356 – 367. doi:10.1016/j.landusepol.2016.06.004.

[15] Mattos, S.H.V.L.d., Vicente, L.E., Vicente, A.K., Júnior, C.B., Piqueira, J.R.C., 2022. Metrics based on information entropy applied to evaluate complexity of landscape patterns. PLOS ONE 17, 1–23. doi:10.1371/journal.pone.0262680.

[16] Mohammadi, A., Karimzadeh, S., Valizadeh Kamran, K., Matsuoka, M., 2020. Extraction of land information, future landscape changes and seismic hazard assessment: A case study of tabriz, iran. Sensors 20, 7010.

[17] Muñoz-Cancino, R., 2023. On the use of multi-sensor digital traces to discover spatio-temporal human behavioral patterns. Ph.D. thesis. University of Basque Country, San Sebastián, Spain.

[18] Muñoz-Cancino, R., Ríos, S.A., Graña, M., 2023. Predicting innovative cities using spatio-temporal activity patterns, in: Hybrid Artificial Intelligent Systems, Springer Nature Switzerland, Cham. pp. 566–576.

[19] Munoz-Cancino, R., Ríos, S.A., Graña, M., 2023. Geotagged digital traces. doi:10.5281/zenodo.7949307.

[20] Muñoz-Cancino, R., Rios, S.A., Goic, M., Graña, M., 2021. Non-intrusive assessment of covid-19 lockdown follow-up and impact using credit card information: Case study in chile. International Journal of Environmental Research and Public Health 18. doi:10.3390/ijerph18115507.

[21] Muñoz-Cancino, R., Ríos, S.A., Graña, M., 2023. Clustering cities over features extracted from multiple virtual sensors measuring micro-level activity patterns allows one to discriminate large-scale city characteristics. Sensors 23. doi:10.3390/s23115165.

[22] Noulas, A., Scellato, S., Mascolo, C., Pontil, M., 2011. Exploiting semantic annotations for clustering geographic areas and users in location-based social networks. Proceedings of the International AAAI Conference on Web and Social Media, Barcelona, Catalonia, Spain, July 21 5, 32–35. doi:10.1609/icwsm.v5i3.14212.

[23] Perola, E., Todorovic, S., Muukkonen, P., Järv, O., 2020. Exploratory visual methods to aggregate origin-destination geodata. Examples and progress in geodata science .

[24] Ríos, S.A., Muñoz, R., 2017. Land use detection with cell phone data using topic models: Case santiago, chile. Computers, Environment and Urban Systems 61, 39–48.

[25] Sapena, M., Ruiz, L.A., 2021. Identifying urban growth patterns through land-use/land-cover spatio-temporal metrics: Simulation and analysis. International Journal of Geographical Information Science 35, 375–396. doi:10.1080/13658816.2020.1817463.

[26] Schafer, V., Truc, G., Badouard, R., Castex, L., Musiani, F., 2019. Paris and nice terrorist attacks: Exploring twitter and web archives. Media, War & Conflict 12, 153–170. doi:10.1177/1750635219839382.

[27] Valls, F., Roca, J., 2021. Visualizing digital traces for sustainable urban management: mapping tourism activity on the virtual public space. Sustainability 13, 3159.

[28] Waller, S.T., Chand, S., Zlojutro, A., Nair, D., Niu, C., Wang, J., Zhang, X., Dixit, V.V., 2021. Rapidex: A novel tool to estimate origin–destination trips using pervasive traffic data. Sustainability 13. doi:10.3390/su132011171.

[29] Wang, D., Gong, J., Chen, L., Zhang, L., Song, Y., Yue, Y., 2012. Spatio-temporal pattern analysis of land use/cover change trajectories in xihe watershed. International Journal of Applied Earth Observation and Geoinformation 14, 12–21. doi:10.1016/j.jag.2011.08.007.

[30] Wang, Q., Taylor, J.E., 2014. Quantifying human mobility perturbation and resilience in hurricane sandy. PLOS ONE 9, 1–5. doi:10.1371/journal.pone.0112608.

[31] Willis, A., Gjersoe, N., Havard, C., Kerridge, J., Kukla, R., 2004. Human movement behaviour in urban spaces: Implications for the design and modelling of effective pedestrian environments. Environment and Planning B: Planning and Design 31, 805–828.

[32] Yan, Y., Schultz, M., Zipf, A., 2019. An exploratory analysis of usability of flickr tags for land use/land cover attribution. Geo-Spatial Information Science 22, 12–22.

[33] Yu, M., Yang, C., Li, Y., 2018. Big data in natural disaster management: A review. Geosciences 8. doi:10.3390/geosciences8050165.

[34] Zhao, K., Tarkoma, S., Liu, S., Vo, H., 2016. Urban human mobility data mining: An overview, in: 2016 IEEE International Conference on Big Data (Big Data), IEEE. pp. 1911–1920.