Establishing Patterns of the Urban Transport Flows on Clustering Analysis*

Vitaliy Pavlyshyn[1,†], Oleksander Ryzhanskyi[1,†], Eduard Manziuk[1,†], Pavlo Radiuk[1,*,†], Oleksander Barmak[1,†] and Iurii Krak[2,3,†]

[1] Khmelnytskyi National University, 11, Institutes str., Khmelnytskyi, 29016, Ukraine
[2] Taras Shevchenko National University of Kyiv, 64/13, Volodymyrska str., Kyiv, 01601, Ukraine
[3] Glushkov Cybernetics Institute, 40, Glushkov Ave., Kyiv, 03187, Ukraine

Abstract

The article presents an adaptive method for identifying characteristic traffic modes in the urban transport environment based on cluster analysis. The developed hierarchical model of classification of transport patterns provides analysis at different levels of detail – from local changes at individual intersections to global modes of operation of the entire network. The proposed multidimensional methodology for assessing the similarity of transport states considers average values, variability, dynamics of changes and time dependencies, providing higher classification accuracy than traditional approaches. Adaptive analysis of time windows automatically adjusts the duration of the study interval depending on the dynamics of traffic flow, allowing you to effectively identify both short-term changes and long-term cyclical patterns. The developed hybrid clustering algorithm, integrating HDBSCAN and k-means methods, demonstrates high noise immunity while maintaining computational efficiency. The method's effectiveness was confirmed experimentally on a simulation model of the transport network of the city of Khmelnytskyi, where four basic traffic scenarios were successfully identified. The analysis of the silhouette coefficients showed the advantage of the HDBSCAN method with an index of 0.37 over the k-means with an index of 0.26 at K = 6, which confirms the effectiveness of the automatic determination of the optimal number of clusters. The results create the basis for optimizing urban transport management, improving traffic safety and improving the quality of transport services.

Keywords

Traffic patterns, clustering analysis, urban transportation, adaptive time windows, hierarchical classification

1. Introduction

The rapid development of urban infrastructure and the increasing use of vehicles pose significant challenges for traffic management. Large volumes of data on traffic flows open up opportunities for their analysis and optimization of urban transport systems. Clustering methods are especially promising for identifying hidden patterns and grouping transport modes according to similar characteristics [1, 2]. They allow for analyzing complex interactions that are difficult to identify with traditional methods. The identification of characteristic traffic patterns is complicated by temporal and spatial variations, high data dimensionality and the dynamic nature of urban traffic [3]. Clustering techniques help identify natural groups in transport data, and recent advances in machine learning have expanded their capabilities for analyzing traffic flows.

However, existing approaches have limitations regarding computational efficiency, real-time processing capabilities, and the ability to analyze long-term time patterns. Many studies do not consider long-term trends and the influence of external factors on movement.

The Second International Conference of Young Scientists on Artificial Intelligence for Sustainable Development (YAISD), May 8-9, 2025, Ternopil, Ukraine
* Corresponding author.
† These authors contributed equally.
vitaliy@ualeaders.com (V. Pavlyshyn); alex@ualeaders.com (O. Ryzhanskyi); radiukp@khmnu.edu.ua (P. Radiuk); manziuk.e@khmnu.edu.ua (E. Manziuk); barmako@khmnu.edu.ua (O. Barmak); iurii.krak@knu.ua (I. Krak)
0009-0003-0702-9359 (V. Pavlyshyn); 0009-0000-4664-5195 (O. Ryzhanskyi); 0000-0003-3609-112X (P. Radiuk); 0000-0002-7310-2126 (E. Manziuk); 0000-0003-0739-9678 (O. Barmak); 0000-0002-8043-0785 (I. Krak)

The study aims to overcome these limitations by developing an improved approach to identifying characteristic driving modes. The main scientific contribution of the article is:

- Development of a hierarchical model of classification of transport patterns at different levels of detail.
- Creation of a multidimensional methodology for assessing the similarity of transport states
- Development of a method of adaptive analysis of time windows.
- Creation of a hybrid clustering algorithm with an innovative pattern validation mechanism.

The proposed approach integrates traditional clustering and machine learning methods to analyze traffic patterns and improve urban traffic management effectively.

The remainder of the paper is structured as follows: Sect. 2 reviews the literature on developing clustering methods in transport systems with an analysis of their limitations. Sect. 3 presents the developed adaptive method for identifying characteristic traffic modes with its mathematical formulation. Sect. 4 describes the results of an experimental study on a simulation model of the transport network of the city of Khmelnytskyi. Sect. 5 addresses analysis and discussing the received result with accent on the practical application of the method. Sect. 6 presents the study's conclusions, emphasizing key scientific contributions.

2. Related works

The development of transport systems and information technologies leads to the generation of significant amounts of data on the movement of vehicles, and clustering methods allow you to identify hidden patterns and group objects according to similar characteristics.

In the field of traffic analysis, research [4–6] were focused on the development of methods for classifying traffic conditions. However, there is a need to consider the landscape features of roads, traffic priorities, and other aspects [7, 8]. In research [4], a hybrid method combining K-medoids and spectral clustering is proposed, and in [5] a spatially constrained hierarchical clustering algorithm for traffic forecasting has been developed. Researches [9, 10] presented a Bayesian model of ensemble clustering of Gaussian processes and an improved clustering scheme based on self-learning. Researches [11, 12] proposed methodologies for assessing traffic conditions based on GPS data. However, these studies are limited to a short analysis period and do not sufficiently consider external factors.

In transport networks and communication systems, works [13-16] proposed various clustering approaches to improve VANET networks, including routing protocols and Harris Hawks optimization. Research [17] presented an approach to identify patterns of mobility, and in [18] the clustering of data on road accidents has been studied. The combination of structural ontology alignment with deep explanatory learning through transition matrices reveals patterns of urban traffic flows during clustering [19, 20]. Works [21, 22] considered hierarchical clustering in transport systems, but problems with scalability and data security were identified.

To analyze the traffic trajectories, studies [23–25] focused on clustering the trajectories of different vehicles, and in research [26] an overview of the clustering of public transport users was carried out. Works [27–29] investigated the application of deep learning for clustering trajectories. The main limitations include low accuracy in measuring trajectory similarities and parameter sensitivity. Based on research analysis [30–32], several key areas for future research have been identified: the development of more effective methods for assessing the similarity of objects, the creation of adaptive clustering algorithms for real time, the improvement of visualization of results, and the development of methods for assessing the quality of clustering. Thus, the purpose of this study is to develop a comprehensive method for identifying characteristic traffic modes in the urban transport environment to increase the efficiency of urban traffic flow management

3. Adaptive method for identifying characteristic driving modes

Consider the urban transport network, presented in the form of an oriented graph

$$G = (V, E), \tag{1}$$

where V is the set of intersections, E is the set of road segments connecting them.

The state of the transport network at any given time t can be represented as a multidimensional vector

$$S(t) = \{s_1(t), s_2(t), \ldots, s_n(t)\}, \tag{2}$$

where $s_i(t)$ represents the state of the intersection i at time t and is also a vector

$$s_i(t) = \{q_{i,1}(t), q_{i,2}(t), \ldots, q_{i,m_i}(t)\}, \tag{3}$$

where $q_{i,j}(t)$ – represents the length of the queue in the direction j at the intersection i at time t, m_i is the number of possible directions of movement at intersection i.

For the time interval $[t_0, t_N]$, we get the sequence of network states $SSN = \{S(t_0), S(t_1), \ldots, S(t_N)\}$. The segmentation function φ: $SSN \longrightarrow SW$ maps the output time series to a sequence of windows: $SW = \{W_1, W_2, \ldots, W_k\}$,

$$W_k = \{S(t) \mid t \in [t_0 + (k-1)\Delta t, \mathrm{t}_0 + k\Delta t]\}. \tag{4}$$

For each window $\mathrm{W_k}$, we calculate the vector of characteristics

$$Ftrk = \{\mu_k, \sigma_k, \delta_k, \tau_k\}, \tag{5}$$

where μ_k is the average state of traffic, σ_k is the standard deviation, δ_k is the rate of change of flows, τ_k is the time dependencies.

The measure of similarity between windows is defined as follows

$$\mathrm{sim}(W_i, W_j) = \exp - \frac{\|Ftr(W_i) - Ftr(W_j)\|}{2\alpha^2}, \tag{6}$$

where $Ftr(W_i)$ and $Ftr(W_j)$ are the vectors of the characteristics of the W_i and W_j windows, α is the scaling parameter.

A sequence of windows is considered a pattern if for all $j \in [\mathrm{i}, \mathrm{i}+k-1]$

$$\mathrm{sim}(W_{\mathrm{j}}, W_{j+1}) > \theta_{\mathrm{threshold}}. \tag{7}$$

Pattern stability is defined as

$$Stb(Prn) = \min_{j,k \in Prn} \mathrm{sim}(W_j, W_k). \tag{8}$$

The optimal window size is defined as

$$\Delta t = \mathrm{argmin}_{\delta} \{Loss(\delta) + \lambda Reg(\delta)\}. \tag{9}$$

A pattern is considered defined if

$$Val(Prn) = Stb(Prn) \cdot Len(Prn) > \theta_{val}. \tag{10}$$

For clustering, we use k-means, where the algorithm minimizes

$$FA = \sum_{k=1}^{K} \sum_{i \in C_k} \|Frt_i - cnt_k\| \tag{11}$$

and HDBSCAN with parameters $\min_{\mathrm{cluster\_size}} = \lceil Nob \cdot scl \rceil$, $\min_{\mathrm{samples}} = \lceil \min_{\mathrm{cluster\_size}} \cdot \beta \rceil$, ε = median (KNNdist)$\cdot\gamma$.

The compactness of the cluster is calculated using the formula

$$Cmp(C_k) = \frac{1}{|C_k|(|C_k| - 1)} \sum_{i,j} sim(W_i, W_j). \tag{12}$$

The formalized algorithm for determining patterns includes data collection and preparation, formation of time windows, clustering of HDBSCAN and k-means, pattern detection, and formation of movement modes.

For a visual representation of the general stages of the method for determining the patterns of transport flows, a diagram (Figure 1) has been developed. It demonstrates the relationship between different stages of data processing. The diagram illustrates the full cycle from obtaining input data on the state of the transport network to the formation of traffic modes and the matrix of transitions between them, visualizing the main components of the proposed algorithm.

The presented diagram illustrates the complex structure of the method for determining transport patterns, where the input data (state of intersections, queue lengths, timestamps, network topology) are sequentially transformed through the stages of formation of state vectors with characteristics {μ, σ, δ, τ}, formation of time windows, parallel clustering by k-means and HDBSCAN methods with subsequent selection of the optimal result, and detection of patterns according to the criteria of length, continuity, stability and belonging to the cluster.

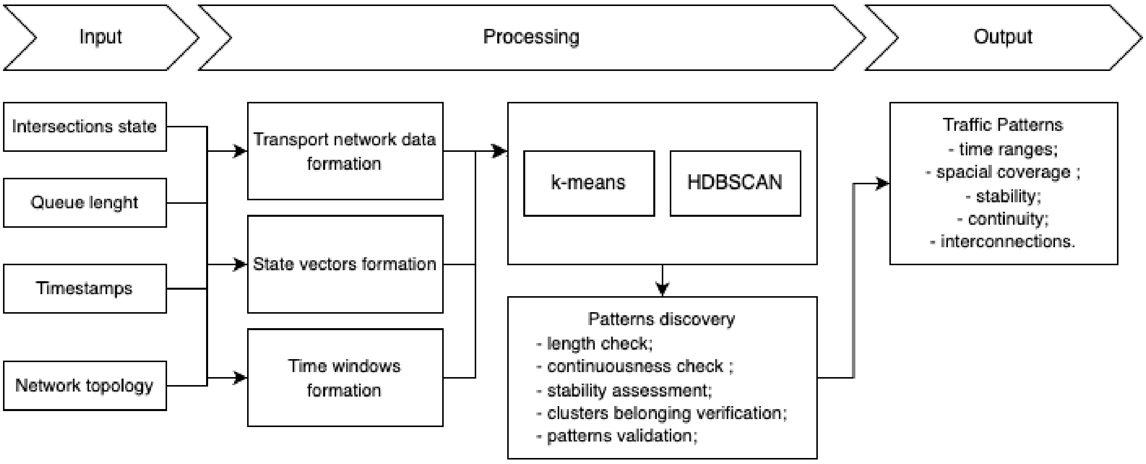


Figure 1: Stages of the method for determining traffic flow patterns.

The analysis results are traffic modes with their characteristics (time ranges, spatial coverage, stability, continuity, relationships) and a transition matrix that simulates dynamic changes between modes, allowing prediction of future states of the transport network, identifying typical sequences of modes and identifying anomalous situations.

4. Results

To validate the proposed method for identifying characteristic traffic modes in the urban environment, an experimental study was conducted using a simulation model of the transport network of the city of Khmelnytskyi. The experiment aimed to confirm the effectiveness of the developed method for detecting and classifying transport patterns in conditions close to real ones.

The experiment's methodology was based on generating and analyzing traffic flows according to various traffic scenarios reflecting typical situations in the urban environment. This approach made it possible to assess the ability of the algorithm to recognize stable patterns in conditions of variability of transport data.

Although the silhouette method provides a mathematical assessment of clustering quality, the final selection of the number of K clusters in transport analysis often requires a combined approach. The determination of K occurs empirically, taking into account the expert assessment of transport engineers regarding typical traffic modes in a particular transport network, the analysis of historical data on the characteristic states of the transport system, the specifics of the area under study, including the size of the city, types of roads and population density, as well as seasonal and daily cycles of transport activity. At the same time, experts consider a complex of interrelated factors. Typical driving modes such as night, morning peak, day and evening peak play a significant role. An important place is occupied by specific system conditions associated with mass events, holiday periods and repair work. Different levels of congestion, from low to critical, as well as weather conditions and their impact on driving modes, have a significant impact.

The empirical approach allows you to validate and, if necessary, adjust the results of the mathematical estimation of the optimal number of clusters, providing a more practically significant clustering of transport patterns.

A temporal sequence of traffic flows was created for the experiment, consisting of different traffic scenarios typical of the urban environment. Each scenario reflected a specific mode of movement of vehicles with variations in the intensity of flows:

1. Morning scenario (0:00-1:00, 5:20-6:40) – characterized by the movement of vehicles toward the city center and the clothing market, which is typical for the morning rush hour.

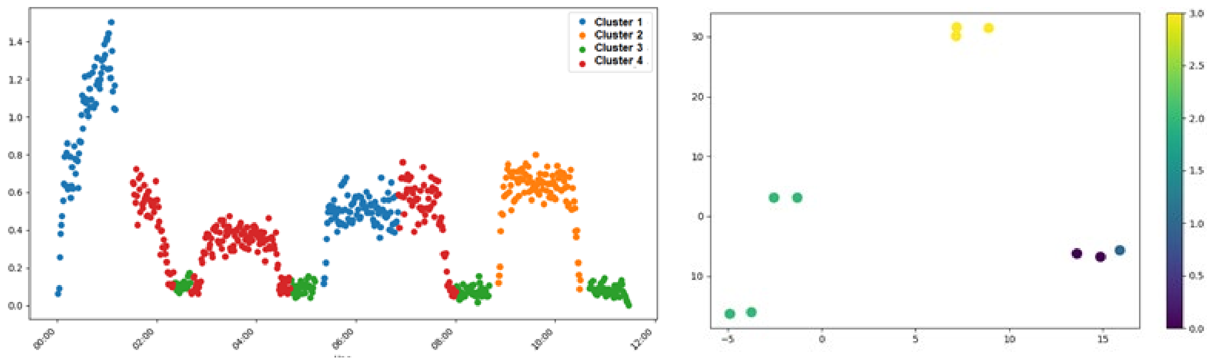
2. Evening scenario (1:00-2:00, 2:50-4:20, 6:40-7:40) – reflects the movement of vehicles from the city center and the clothing market, which is typical for the evening rush hour, with variations in the intensity of flows.

3. Scenario of the Greceany district (2:00-2:50, 4:20-5:20, 7:40-8:50, 10:20-11:20) – represents the movement of vehicles from the Greceany district to the city center and the clothing market, as well as in the opposite direction, with different intensity.

4. Mixed scenario (8:50-10:20) – combines elements of morning and evening scenarios with reduced traffic intensity in all directions.

Each state of the transport network was presented as a multidimensional vector containing information. The vector representation made it possible to preserve the complete structure of transport data and the relationships between different directions of movement. Two clustering methods–HDBSCAN and k-means–were applied to identify characteristic modes of motion, according to the algorithm described in the previous section.

This made it possible to compare the effectiveness of different approaches to detecting transport patterns and confirm the reliability of the proposed method. The use of the HDBSCAN method made it possible to automatically determine the optimal number of clusters without first specifying this parameter, which is a significant advantage in the analysis of dynamic transport data. The results of clustering are presented in Figure 2.



(a) (b)

Figure 2: (a) Clustering with HDBSCAN Transport Flow Pattern Detection, (b) Visualization of HDBSCAN clustering results using the UMAP method

As can be seen from Figure 2a, the HDBSCAN method successfully identified four clearly separated clusters corresponding to the main motion scenarios:

- Cluster 1, which corresponds to the morning scenario.
- Cluster 2, which corresponds to the evening scenario.
- Cluster 3, which corresponds to the scenario of the Greceani district.
- Cluster 4, which corresponds to a mixed scenario.

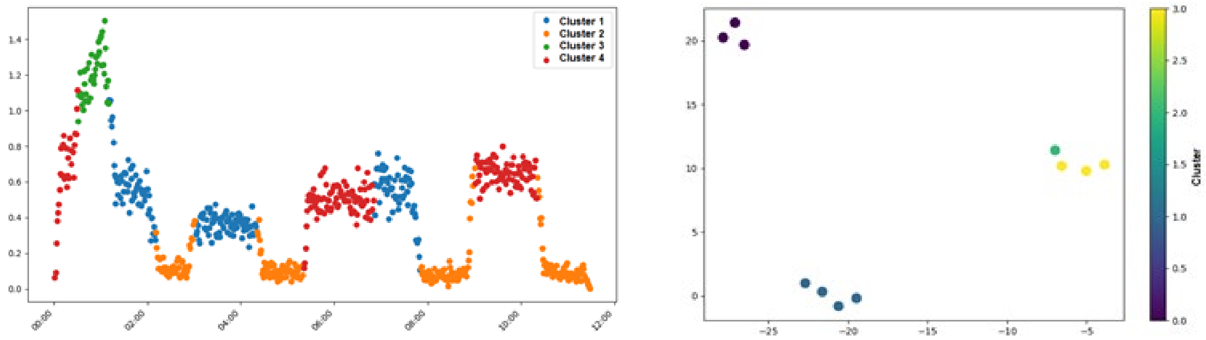
An important feature of the obtained results is the absence of intersections between clusters in the time dimension, which indicates high classification accuracy and clear differentiation of different modes of motion. This confirms the effectiveness of the proposed method for detecting characteristic transport patterns.

The UMAP method was used to visualize multidimensional clustering data, which made it possible to display the results in two-dimensional space (Figure 2b). UMAP visualization demonstrates a clear separation of clusters in two-dimensional space, further confirming the effectiveness of the HDBSCAN method for identifying transport patterns. It is important to note that UMAP displays data not by time component but by the similarity of internal characteristics of traffic flows, which allows you to identify hidden patterns in multidimensional data.

The k-means method was also applied to validate the results and benchmarking with a predetermined number of clusters K=4, corresponding to the number of scenarios in the experimental data. The results of clustering are presented in Figure 3a.

Comparison of k-means results with HDBSCAN results shows high consistency between the two methods. K-Means also successfully identified four clusters that generally correspond to the main motion scenarios identified in the experimental data. Visualization of k-means results using the

UMAP method (Figure 3b) also demonstrates a clear separation of clusters in two-dimensional space.



(a) (b)

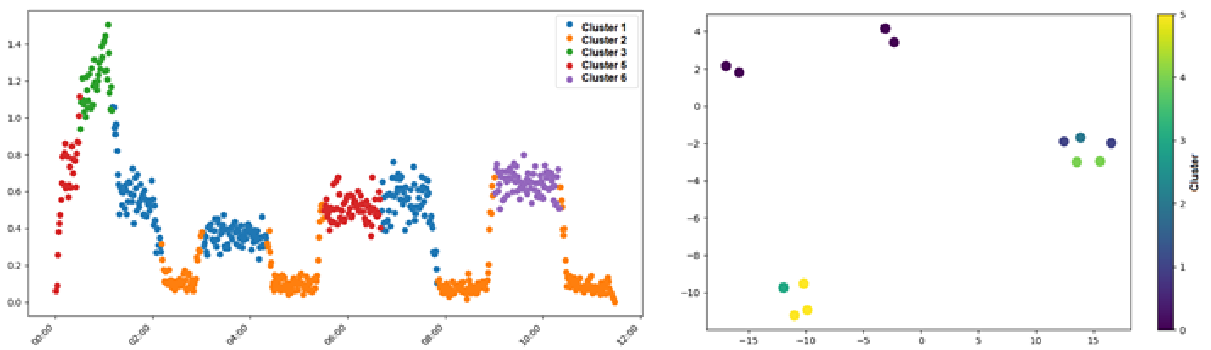
Figure 3: (a) Results of clustering of transport flows using the k-means method (K = 4), (b) Visualization of k-means clustering results (K=4) using the UMAP method

Increasing the number of clusters to K = 6 resulted in a more detailed but less consistent classification with the initial scenarios. Some scenarios have been divided into subcategories, which can be useful for more nuanced analysis but complicates the overall interpretation of the results. This confirms the advantage of the HDBSCAN method, which automatically determines the optimal number of clusters based on the data structure.

Thus, the application of the HDBSCAN method made it possible to identify the optimal number of clusters successfully, in our case K = 4), corresponding to the main movement scenarios in the experimental data, without the need to pre-set this parameter.

This result confirms the method's effectiveness for analyzing traffic flows with a structure unknown in advance and demonstrates the need to involve expert assessment for cluster validation. An important aspect of the study was the high consistency of the results obtained using two different clustering methods. Such consistency confirms the stability of the identified patterns and the reliability of the proposed approach to identifying transport patterns.

Additionally, an experiment was carried out using the k-means method with K = 6 to investigate the possibility of a more detailed classification of transport modes (Figure 6).



(a) (b)

Figure 4: (a) Results of clustering transport flows using the k-means method (K = 6), (b) Visualization of k-means clustering results (K = 6) using the UMAP method

In the study, both methods demonstrated a clear separation of clusters in both the temporal dimension and the characteristic space, which was confirmed by UMAP imaging. This indicates the high quality of the clustering carried out and the ability of the proposed method to identify different modes of movement effectively. An experiment using k-means at K = 6 demonstrated that increasing the number of clusters can lead to more detailed but potentially over-classification, making the results difficult to interpret. This fact emphasizes the importance of optimal selection of the number of clusters, which is one of the key advantages of the HDBSCAN method.

The traffic modes detected through clustering demonstrate a clear correspondence to typical transport scenarios in the urban environment, such as morning rush time, evening rush time and local traffic modes in selected areas of the city. Each cluster is characterized by a unique set of transport characteristics, effectively distinguishing between different transport network states.

5. Discussion

The experiment's results confirm the effectiveness of the proposed method for identifying characteristic modes of movement. Multivariate similarity estimation, considering the mean values of the flows, their variability, dynamics of changes and time dependencies, provided higher classification accuracy than traditional approaches, which are often limited to analyzing only one or two parameters. The adaptive mechanism for selecting time windows has demonstrated high flexibility when working with traffic flows of different dynamics. During the experiment, the optimal window size for the morning and evening scenarios was 15–20 minutes, and for more stable periods, it increased to 30–40 minutes.

A comparative analysis of the HDBSCAN and k-means methods revealed the former's advantage in automatically determining the optimal number of clusters and in higher noise immunity. At the same time, k-means demonstrated better computing efficiency. Increasing the number of clusters in the k-means method from 4 to 6 resulted in a more detailed but potentially over-classification, making the results difficult to interpret. For the HDBSCAN algorithm, the average silhouette coefficient is 0.37. Clusters 2 and 3 demonstrate the highest values (0.2–0.9), while clusters 1 and 4 have lower values (0.0–0.6) with some negative points. The average silhouette coefficient for the K-means with 6 clusters is lower – 0.26. Only clusters 1 and 2 show relatively high values (0.1–0.7), while the remaining clusters (3–6) have mostly low values (0.0–0.3), indicating their potential redundancy. HDBSCAN demonstrates better clustering quality due to higher silhouette coefficients and clearer cluster separation.

Visualization of the results by the UMAP method confirmed the effectiveness of the chosen approach to presenting transport data, demonstrating a clear separation of clusters in two-dimensional space. The method has certain limitations in terms of its application in cases of changes in the transportation network, such as temporary road closures or changes in traffic patterns. However, it applies to special cases of short-term impact. From a practical point of view, the developed method can potentially optimize traffic light regulation, strategic planning of transport infrastructure and forecasting the transport situation to prevent congestion proactively.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly in order to: Grammar and spelling check. After using this tool/service, the authors reviewed and edited the content as needed and takes full responsibility for the publication's content.

6. Conclusions

As a result of the study, a method for identifying characteristic traffic modes in the urban transport environment was developed and experimentally validated. The main scientific contribution was developing a hierarchical model for classifying transport patterns with a multidimensional assessment of the similarity of states, which provides analysis at different levels of detail – from local changes at intersections to global modes of network functioning.

The key components of the method are an adaptive time-window mechanism that automatically adjusts the duration of the study interval depending on the dynamics of the transport flow, and a hybrid clustering algorithm that integrates HDBSCAN and k-means methods with an innovative pattern validation mechanism. This provides high resistance to noise and anomalies while maintaining computational efficiency. The analysis of the silhouette coefficients demonstrates the superiority of the HDBSCAN algorithm with a score of 0.37 over K-means with a score of 0.26, confirming the feasibility of automatically determining the optimal number of clusters for effective classification of traffic patterns.

The method's effectiveness was confirmed experimentally on a simulation model of the transport network of the city of Khmelnytskyi, where four basic traffic scenarios were successfully identified – morning, evening, specific for the Grechany district and mixed. The results create a methodological basis for optimizing urban transport management, improving traffic safety and improving the quality of transport services.

Acknowledgements

This research was supported by the U_CAN – "Towards carbon neutrality of Ukrainian cities" project, under Grant Agreement No. 101148374, funded by the Horizon Europe program.

References

[1] Ž. Majstorović, L. Tišljarić, E. Ivanjko, and T. Carić, "Urban traffic signal control under mixed traffic flows: Literature review," Applied Sciences, vol. 13, no. 7, Art. no. 7, Jan. 2023, doi: 10.3390/app13074484.

[2] M. Chaudhry, I. Shafi, M. Mahnoor, D. L. R. Vargas, E. B. Thompson, and I. Ashraf, "A systematic literature review on identifying patterns using unsupervised clustering algorithms: A data mining perspective," Symmetry, vol. 15, no. 9, Art. no. 9, Sep. 2023, doi: 10.3390/sym15091679.

[3] I. Mavlutova, D. Atstaja, J. Grasis, J. Kuzmina, I. Uvarova, and D. Roga, "Urban transportation concept and sustainable urban mobility in smart cities: A review," Energies, vol. 16, no. 8, Art. no. 8, Jan. 2023, doi: 10.3390/en16083585.

[4] Q. Shang, Y. Yu, and T. Xie, "A hybrid method for traffic state classification using k-medoids clustering and self-tuning spectral clustering," Sustainability, vol. 14, no. 17, Art. no. 17, Jan. 2022, doi: 10.3390/su141711068.

[5] K. Kim, "Spatial contiguity-constrained hierarchical clustering for traffic prediction in bike sharing systems," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 6, pp. 5754–5764, Jun. 2022, doi: 10.1109/TITS.2021.3057596.

[6] M. Zhang, "Decision support approach for integrated maintenance program of urban rail transit," *Int. J. Comput.*, vol. 16, no. 3, pp. 143–151. doi: 10.47839/ijc.16.3.897.

[7] O. Ryzhanskyi, E. Manziuk, O. Barmak, I. Krak, and N. Bačanin, "An approach to optimizing CO2 emissions in traffic control via reinforcement learning," CEUR-WS, vol. 3675, pp. 137–155, 2024, URL: https://ceur-ws.org/Vol-3675/paper10.pdf.

[8] V. Pavlyshyn, E. Manziuk, O. Barmak, I. Krak, and R. Damasevicius, "Modeling environment intelligent transport system for eco-friendly urban mobility," CEUR-WS, vol. 3675, pp. 118–136, 2024, URL: https://ceur-ws.org/Vol-3675/paper10.pdf.

[9] L. Khrissi, N. El Akkad, H. Satori, and K. Satori, "A performant clustering approach based on an improved sine cosine algorithm," *Int. J. Comput.*, vol. 21, no. 2. pp. 159–168, doi: 10.47839/ijc.21.2.2584.

[10] Z. Zhu, M. Xu, J. Ke, H. Yang, and X. Chen, "A Bayesian clustering ensemble Gaussian process model for network-wide traffic flow clustering and prediction," *Transp. Res. Part C: Emerg. Technol.*, vol. 148, p. 104032, Mar. 2023. doi: 10.1016/j.trc.2023.104032.

[11] U. Ahmed, R. Tu, J. Xu, G. Amirjamshidi, M. Hatzopoulou, and M. J. Roorda, "GPS-Based Traffic Conditions Classification Using Machine Learning Approaches," Transportation Research Record, vol. 2677, no. 2, pp. 1445–1454, Feb. 2023, doi: 10.1177/03611981221111370.

[12] K. Mohammed Almatar, "Traffic congestion patterns in the urban road network: (Dammam metropolitan area)," Ain Shams Engineering Journal, vol. 14, no. 3, pp. 1–9, Apr. 2023, doi: 10.1016/j.asej.2022.101886.

[13] O. Kyrychenko, S. Ostapov, and I. Malyk, "Cluster analysis of information in complex networks," *Int. J. Comput.*, vol. 22, no. 4, pp. 515–523. doi: 10.47839/ijc.22.4.3360.

[14] A. Ali, F. Aadil, M. F. Khan, M. Maqsood, and S. Lim, "Harris hawks optimization-based clustering algorithm for vehicular ad-hoc networks," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–20, 2023. Accessed: May 2, 2025. doi: 10.1109/TITS.2023.3257484.

[15] V. Turchenko, L. Grandinetti, and A. Sachenko, “Parallel batch pattern training of neural networks on computational clusters,” in *2012 Int. Conf. High Perform. Comput. & Simul. (HPCS)*, Madrid, Spain, Jul. 2–6, 2012. New York, NY, USA: IEEE, 2012, pp. 202–208. doi: 10.1109/hpcsim.2012.6266912.

[16] O. Barmak, I. Krak, and E. Manziuk, Diversity as the basis for effective clustering-based classification, in CEUR Workshop Proceedings, Odessa, Ukraine: CEUR, Sep. 2020, pp. 53–67. Accessed: Jul. 24, 2021. URL: http://ceur-ws.org/Vol-2711/paper5.pdf

[17] F. Neves, A. C. Finamore, S. C. Madeira, and R. Henriques, “Mining actionable patterns of road mobility from heterogeneous traffic data using biclustering,” Trans. Intell. Transport. Sys., vol. 23, no. 7, pp. 6430–6445, 2022, doi: 10.1109/TITS.2021.3057240.

[18] E. Esenturk, A. G. Wallace, S. Khastgir, and P. Jennings, “Identification of traffic accident patterns via cluster analysis and test scenario development for autonomous vehicles,” IEEE Access, vol. 10, pp. 6660–6675, 2022, doi: 10.1109/ACCESS.2021.3140052.

[19] P. Radiuk, O. Barmak, E. Manziuk, and I. Krak, “Explainable deep learning: A visual analytics approach with transition matrices,” Mathematics, vol. 12, no. 7, pp. 1–32, 2024, doi: 10.3390/math12071024.

[20] E. Manziuk, I. Krak, O. Barmak, O. Mazurets, V. Kuznetsov, and O. Pylypiak, Structural alignment method of conceptual categories of ontology and formalized domain, presented at the CEUR Workshop Proceedings, 2021, pp. 11–22. URL: http://ceur-ws.org/Vol-3003/paper2.pdf

[21] B. Ghimire and D. B. Rawat, “Dynamic clustering in IoV using behavioral parameters and contention window adaptation,” IEEE Transactions on Vehicular Technology, vol. 71, no. 2, pp. 2031–2040, Feb. 2022, doi: 10.1109/TVT.2021.3134965.

[22] A. K. Sangaiah, A. Javadpour, F. Ja’fari, W. Zhang, and S. M. Khaniabadi, “Hierarchical clustering based on dendrogram in sustainable transportation systems,” IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 12, pp. 15724–15739, Dec. 2023, doi: 10.1109/TITS.2022.3222789.

[23] T. Bolić, L. Castelli, A. De Lorenzo, and F. Vascotto, “Trajectory clustering for air traffic categorisation,” Aerospace, vol. 9, no. 5, Art. no. 5, May 2022, doi: 10.3390/aerospace9050227.

[24] J. Yang, Y. Liu, L. Ma, and C. Ji, “Maritime traffic flow clustering analysis by density based trajectory clustering with noise,” Ocean Engineering, vol. 249, Apr. 2022, doi: 10.1016/j.oceaneng.2022.111001.

[25] B. Y. Chen et al., “Efficient and scalable DBSCAN framework for clustering continuous trajectories in road networks,” International Journal of Geographical Information Science, vol. 37, no. 8, pp. 1693–1727, Aug. 2023, doi: 10.1080/13658816.2023.2217443.

[26] O. Cats, “Identifying human mobility patterns using smart card data,” Transport Reviews, vol. 44, no. 1, pp. 213–243, Jan. 2024, doi: 10.1080/01441647.2023.2251688.

[27] C. Wang et al., “A deep trajectory clustering method based on sequence-to-sequence autoencoder model,” Transactions in GIS, vol. 26, no. 4, pp. 1801–1820, 2022, doi: 10.1111/tgis.12905.

[28] C. Zhang, W. Huang, T. Niu, Z. Liu, G. Li, and D. Cao, “Review of clustering technology and its application in coordinating vehicle subsystems,” Automot. Innov., vol. 6, no. 1, pp. 89–115, Feb. 2023, doi: 10.1007/s42154-022-00205-0.

[29] Y. Zhang, X. Yuan, M. Li, G. Zhao, and H. Wang, “Multi-density adaptive trajectory clustering algorithm for ships based on AIS data,” IEEE Access, vol. PP, pp. 108198–108210, Jan. 2023, doi: 10.1109/ACCESS.2023.3321270.

[30] T. A. Akinosho, E. Tabane, and Z. Wang, “A comparative analysis of data stream clustering algorithms,” *Int. J. Comput.*, vol. 22, no. 4, pp. 439–446. doi: 10.47839/ijc.22.4.3350.

[31] A. Azab, M. Khasawneh, S. Alrabaee, K.-K. R. Choo, and M. Sarsour, “Network traffic classification: Techniques, datasets, and challenges,” Digital Communications and Networks, vol. 10, no. 3, pp. 676–692, Jun. 2024, doi: 10.1016/j.dcan.2022.09.009.

[32] C. Vishnu, V. Abhinav, D. Roy, C. K. Mohan, and Ch. S. Babu, “Improving multi-agent trajectory prediction using traffic states on interactive driving scenarios,” IEEE Robotics and Automation Letters, vol. 8, no. 5, pp. 2708–2715, May 2023, doi: 10.1109/LRA.2023.3258685.