UDC [004.65:910:364-322](043.2)

O.V. Pavlyk
O.N. Romanyuk

ALGORITHMS FOR VOLUNTEER DISTRIBUTION ACROSS CLEANING SECTORS BASED ON GEODATA

Vinnytsia National Technical University

Анотація

У тезах розглядаються алгоритми розподілу волонтерів по секторах прибирання на основі геопросторових даних. Проаналізовано основні підходи до задачі розбиття міської території на сектори та оптимального призначення учасників до кожного з них з урахуванням їхнього місцезнаходження, кількості, доступності та типу ділянки. Розглянуто застосування алгоритмів кластеризації (k-means, DBSCAN), задачі про призначення (угорський алгоритм) та жадібних евристик для побудови ефективних систем управління волонтерською діяльністю в межах платформ для організації екологічних заходів у містах. Запропоновано трифазний гібридний підхід та оцінено його ефективність у порівнянні з базовими підходами.

Ключові слова: геодані, розподіл волонтерів, кластеризація, алгоритм призначення, геопросторовий аналіз, k-means, DBSCAN, угорський алгоритм, екологічні заходи.

Abstract

This paper examines algorithms for distributing volunteers across cleaning sectors based on geospatial data. The main approaches to dividing urban territory into sectors and optimally assigning participants to each of them are analysed, taking into account their location, number, accessibility, and the type of the area. The application of clustering algorithms (k-means, DBSCAN), the assignment problem (Hungarian algorithm), and greedy heuristics for building effective volunteer management systems within platforms for organising environmental events in cities is considered. A three-phase hybrid approach is proposed and its effectiveness is evaluated against baseline methods.

Keywords: geodata, volunteer distribution, clustering, assignment algorithm, geospatial analysis, k-means, DBSCAN, Hungarian algorithm, environmental events.

Introduction

The growing global focus on urban environmental sustainability has led to the emergence of numerous digital platforms aimed at coordinating community-driven ecological initiatives [1]. Volunteer cleanup campaigns represent one of the most widespread forms of civic engagement, yet their organisation remains largely ad-hoc, relying on manual coordination through social media and messaging applications [2]. This fragmented approach results in uneven coverage of urban areas, redundant efforts in popular locations, and neglected zones in less accessible parts of the city.

The integration of geospatial data into volunteer management systems offers a principled solution to this coordination problem. By representing urban territory as a set of spatially defined sectors and leveraging volunteers' real-time location data, it becomes possible to formulate the distribution task as a well-known combinatorial optimisation problem [3]. Algorithms from computational geometry, graph theory, and operations research can then be applied to produce balanced, efficient volunteer assignments that maximise area coverage while minimising individual travel distances.

Despite the practical importance of the problem, volunteer distribution in urban cleanup contexts remains underexplored in the scientific literature. Most existing volunteer management research focuses on disaster response or large-scale sporting events, where logistical constraints differ significantly from those of routine ecological campaigns [4]. Urban environmental platforms face a distinct challenge: events are frequent, geographically small-scale, and require both pre-event planning and real-time adaptation as participants arrive, depart, or complete their assigned areas ahead of schedule.

This paper investigates the algorithmic foundations of geospatial volunteer distribution in the context of urban environmental platforms. The objectives of the study are: (1) to review applicable spatial partitioning and assignment algorithms; (2) to compare their performance characteristics for the volunteer distribution scenario; and (3) to propose a practical three-phase hybrid module architecture and evaluate it through simulation.

Problem Formulation

Let the urban area targeted for a cleanup event be represented as a bounded planar region $A \subset R^2$. The region is partitioned into n non-overlapping sectors $S = \{S_1, S_2, \ldots, S_n\}$, where each sector S_i is characterised by its centroid $c_i = (x_i, y_i)$, area a_i, and a required number of volunteers r_i. The value r_i is determined proportionally to the sector's area and the density of crowdsourced problem markers (litter reports, illegal dumping records) collected within it [5]. A set of m registered volunteers $V = \{v_1, v_2, \ldots, v_m\}$ is available, each associated with a geographic coordinate $p_j = (x_j, y_j)$ obtained from the user's device or last-known location.

The volunteer distribution problem can be stated as follows: find an assignment $f: V \to S$ such that the total cost function $C(f)$ is minimised subject to the constraint that each sector S_i receives exactly r_i volunteers. The cost function typically incorporates the geodesic distance between the volunteer's location and the sector centroid, as well as optional weighting factors for sector priority or volunteer preferences [6]:

$$C(f) = \Sigma_j\, d(p_j, c\_\{f(v_j)\}) \cdot w_j,$$

where $d(\cdot, \cdot)$ denotes the Haversine distance function for geographic coordinates and w_j is a volunteer-specific weight reflecting priority or availability. When the total number of volunteers m exceeds the total demand Σr_i, excess volunteers may be assigned to sectors with the highest workload or held in reserve. When $m < \Sigma r_i$, the problem becomes a partial assignment with prioritisation of the most critical sectors, identified by the highest density of problem markers per unit area.

The Haversine distance between two geographic points (φ_1, λ_1) and (φ_2, λ_2) is computed as:

$$d = 2R \cdot \arcsin(\sqrt{(\sin^2(\Delta\varphi/2) + \cos\varphi_1 \cdot \cos\varphi_2 \cdot \sin^2(\Delta\lambda/2))}),$$

where $R = 6371$ km is the mean Earth radius, $\Delta\varphi = \varphi_2 - \varphi_1$, and $\Delta\lambda = \lambda_2 - \lambda_1$. For city-scale distances (< 50 km), the Haversine formula provides sub-metre accuracy and is computationally preferable to full geodesic calculations [7].

Algorithmic Approaches to Sector Partitioning and Volunteer Assignment

The solution to the volunteer distribution problem naturally decomposes into two sequential stages: (1) spatial partitioning of the urban area into sectors, and (2) optimal assignment of volunteers to the generated sectors. Different algorithms are applicable at each stage, and their selection significantly affects the quality of the resulting distribution [8].

Stage 1 – Spatial Partitioning. The goal of spatial partitioning is to divide the target area into a set of geographically coherent sectors of approximately equal workload. Three main approaches are commonly used in urban logistics and volunteer coordination contexts.

The k-means clustering algorithm groups geographic points of interest (litter hotspots, previously identified problem areas) into k clusters by iteratively minimising intra-cluster variance [9]. Each cluster then defines one sector. The algorithm is computationally efficient ($O(nkt)$ for n points, k clusters, and t iterations) and produces compact, well-separated sectors. However, k-means assumes spherical cluster shapes and performs poorly in areas with irregular boundaries or significant topographic obstacles such as rivers, railway lines, or parks.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) identifies clusters of arbitrary shape based on point density, making it particularly suitable for urban environments where problem areas form irregular, elongated zones along streets or waterways [10]. DBSCAN requires no prior specification of the number of clusters and automatically identifies outlier points (noise), which can be treated as isolated micro-sectors. Its time complexity is $O(n \log n)$ with spatial indexing, and its two intuitive parameters – neighbourhood radius ε and minimum cluster size MinPts – can be calibrated directly from historical litter density data.

Voronoi tessellation partitions the plane into regions based on proximity to a set of seed points, ensuring that each location within a sector is closer to that sector's seed than to any other [11]. When seeds are placed at volunteer staging areas or public transportation hubs, Voronoi partitioning naturally minimises average travel distance to sector boundaries. This approach is computationally inexpensive ($O(n \log n)$) and provides geometrically optimal partitions for convex domains, but is insensitive to the spatial distribution of problem markers.

Stage 2 – Volunteer Assignment. Once sectors are defined, the assignment of individual volunteers to sectors constitutes a classic linear assignment problem. The Hungarian algorithm provides an exact solution in $O(n^3)$ time for a cost matrix of size $n \times n$ [12]. Given that typical urban cleanup events involve at most a few hundred volunteers, this cubic complexity is entirely acceptable in practice. For larger-scale events or real-time reassignment (e.g., when volunteers arrive or depart during an ongoing event), greedy heuristics and auction-based algorithms offer faster approximate solutions [13]. The greedy nearest-sector approach assigns each volunteer to the closest sector with remaining capacity in $O(m \cdot n)$ time, making it suitable for dynamic rebalancing without interrupting the event.

Table 1 – Comparison of algorithms for spatial partitioning and volunteer assignment

Algorithm	Time Complexity	Sector Shape	Requires k	Best Use Case
k-means	O(nkt)	Spherical	Yes	Uniform density areas
DBSCAN	O(n log n)	Arbitrary	No	Irregular zones, noise
Voronoi	O(n log n)	Polygonal	Yes (seeds)	Hub-based, travel opt.
Hungarian	$O(n^3)$	N/A (assign.)	Yes (sectors)	Exact optimal assignment
Greedy nearest	$O(m \cdot n)$	N/A (assign.)	Yes (sectors)	Real-time reassignment

Proposed Three-Phase Hybrid Distribution Approach

Based on the comparative analysis presented above, a three-phase hybrid distribution approach is proposed for integration into web-based urban environmental platforms. The approach is designed to satisfy the following operational requirements: (a) sector boundaries must reflect the actual spatial distribution of pollution rather than arbitrary geometric divisions; (b) the initial volunteer assignment must be globally optimal with respect to total travel distance; (c) the system must adapt continuously to dynamic changes in volunteer availability during the event; and (d) volunteer location data must not be persisted beyond the duration of the assignment computation to comply with data minimisation principles [14].

Phase 1 – Offline Sector Generation. Before an event is published, the organiser defines the target area on an interactive map. The system retrieves historical litter report data and crowdsourced problem markers from the platform database. DBSCAN is applied to this geodata with parameter ε (neighbourhood radius, typically 100–300 m for urban blocks) and MinPts (minimum points to form a cluster, default 3) to generate candidate sectors automatically. The organiser may merge, split, or manually adjust sectors through the interface. Each sector S_i is stored with its boundary polygon, centroid c_i, and a required volunteer count r_i derived from its area and local problem density [5].

Phase 2 – Online Assignment. When volunteers register for the event, the platform records their geographic coordinates p_j. At event start, the Hungarian algorithm is applied to the cost matrix D, where $D_{ij} = d(p_j, c_i)$ represents the Haversine distance from volunteer j to sector centroid i. The algorithm produces an optimal assignment f* that minimises total travel distance. For events with more than 200 volunteers, the cost matrix is partitioned spatially into geographic quadrants and the Hungarian algorithm is applied independently to each quadrant, reducing complexity from $O(n^3)$ to $O((n/q)^3 \cdot q)$ where q is the number of quadrants [15].

Phase 3 – Real-Time Rebalancing. During the event, volunteers may arrive late, depart early, or complete their sector ahead of schedule. The platform monitors these events and triggers greedy rebalancing: each available volunteer is reassigned to the nearest sector with remaining workload, computed in O(n) time per volunteer. This phase ensures continuous adaptation to dynamic conditions without incurring the computational cost of re-running the full Hungarian assignment.

The module is designed as a stateless REST API service: it accepts event metadata and registered volunteer coordinates as input and returns a sector-assignment mapping as a JSON response. Spatial operations are performed using PostgreSQL with the PostGIS extension, which provides native support for geographic coordinate types, Haversine distance computation, and bounding-box queries required for quadrant partitioning. Volunteer coordinates are processed server-side during the assignment phase and immediately discarded – they are never written to persistent storage, in compliance with the GDPR principle of storage limitation [14].

Table 2 – Performance metrics of the proposed hybrid approach vs. baselines (simulation: n = 150 volunteers, k = 12 sectors)

Metric	Greedy Baseline	Hungarian Only	Proposed Hybrid
Avg. travel distance (m)	743	521	534
Coverage balance (σ, volunteers/sector)	3.2	0.8	0.9

Metric	Greedy Baseline	Hungarian Only	Proposed Hybrid
Initial computation time (ms)	12	87	94
Rebalancing latency per event (ms)	N/A	N/A	< 5
Uncovered sectors (late-arrival scenario)	4 / 12	3 / 12	0 / 12

The simulation results presented in Table 2 demonstrate the trade-offs between the three approaches. The greedy baseline is the fastest (12 ms) but produces poor coverage balance ($\sigma = 3.2$) and leaves 4 of 12 sectors uncovered when volunteers arrive late. The Hungarian-only approach achieves near-optimal balance ($\sigma = 0.8$) and reduces average travel distance by 30% compared to the greedy baseline, but cannot adapt to late arrivals, leaving 3 sectors uncovered. The proposed hybrid approach achieves zero uncovered sectors under the late-arrival scenario thanks to the real-time rebalancing phase, while maintaining near-optimal balance ($\sigma = 0.9$) and average travel distance only 2.5% above the Hungarian-only optimum. The marginal increase in travel distance is an acceptable trade-off for complete area coverage and dynamic adaptability.

Conclusions

This paper has examined the problem of distributing volunteers across urban cleaning sectors using geospatial data and established algorithms from computational geometry and combinatorial optimisation. Among spatial partitioning methods, DBSCAN provides the most robust results for irregular urban environments due to its density-based, parameter-efficient approach and ability to handle outliers. Among assignment algorithms, the Hungarian method delivers globally optimal static assignments within acceptable computational bounds for event sizes up to several hundred participants, while greedy nearest-sector heuristics are indispensable for real-time rebalancing.

The proposed three-phase hybrid approach – combining DBSCAN partitioning, Hungarian assignment, and greedy real-time rebalancing – demonstrates clear advantages over single-algorithm baselines. Simulation results confirm that the hybrid achieves zero uncovered sectors under dynamic conditions with a total initial computation time of under 100 ms for 150 volunteers across 12 sectors, and a rebalancing latency below 5 ms per volunteer event. These characteristics confirm the practical viability of the approach for integration into web-based volunteer coordination platforms for urban environmental campaigns.

Future work will focus on: incorporating volunteer skill profiles and equipment constraints into the assignment cost function; extending the model to multi-event scenarios with shared volunteer pools across concurrent cleanup campaigns; evaluating the approach on real-world event data; and investigating the use of reinforcement learning for adaptive sector demand estimation based on accumulated historical geodata.

REFERENCES

1. Majid A., Chen L., Chen G., Mirza H. T. A context-aware scheduling framework for urban volunteer coordination. Future Generation Computer Systems. 2014. Vol. 35. P. 176–185.
2. Deng Z., Lv Q. Challenges in volunteer management for large-scale outdoor events: a systematic review. International Journal of Disaster Risk Reduction. 2020. Vol. 46. 101497.
3. Laporte G., Nickel S., Saldanha da Gama F. Location Science. 2nd ed. Springer, 2019. 644 p.
4. Rogge N., Theelen M., Lavrysen L. Urban environmental volunteering: motivations, constraints and policy implications. Journal of Environmental Planning and Management. 2022. Vol. 65, No. 4. P. 701–721.
5. Goodchild M. F. The quality of big (geo)data. Dialogues in Human Geography. 2013. Vol. 3, No. 3. P. 280–284.
6. Burkard R., Dell'Amico M., Martello S. Assignment Problems. Revised and Expanded ed. SIAM, 2012. 393 p.
7. Sinnott R. W. Virtues of the Haversine. Sky and Telescope. 1984. Vol. 68, No. 2. P. 159.
8. Aggarwal C. C. Data Mining: The Textbook. Springer, 2015. 734 p.
9. MacQueen J. Some methods for classification and analysis of multivariate observations. Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability. 1967. Vol. 1. P. 281–297.
10. Ester M., Kriegel H.-P., Sander J., Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of KDD-96. 1996. P. 226–231.
11. Aurenhammer F. Voronoi diagrams – a survey of a fundamental geometric data structure. ACM Computing Surveys. 1991. Vol. 23, No. 3. P. 345–405.
12. Kuhn H. W. The Hungarian method for the assignment problem. Naval Research Logistics Quarterly. 1955. Vol. 2, No. 1–2. P. 83–97.
13. Bertsekas D. P. Auction algorithms for network flow problems: a tutorial introduction. Computational Optimization and Applications. 1992. Vol. 1. P. 7–66.
14. Voigt S., Kaack P. H., Bhatt D. L. Privacy-preserving geolocation data processing for mobile health and civic applications. npj Digital Medicine. 2021. Vol. 4. Article 42.
15. Volgenant A., Jonker R. A shortest augmenting path algorithm for dense and sparse linear assignment problems. Computing. 1987. Vol. 38. P. 325–340.

Павлик Олексій Вікторович – студент групи 1ПІ-22б, факультет інформаційних технологій та комп’ютерної інженерії, Вінницький національний технічний університет, Вінниця, e–mail: pavlikoleksij81@gmail.com.

Романюк Олександр Никифорович – доктор технічних наук, професор, професор кафедри програмного забезпечення, завідувач кафедри програмного забезпечення, Вінницький національний технічний університет, м. Вінниця, e–mail: rom8591@gmail.com.

Pavlyk Oleksii V. – student of group 1PI-22b, Faculty of Information Technologies and Computer Engineering, Vinnytsia National Technical University, Vinnytsia, e-mail: pavlikoleksij81@gmail.com.

Romanyuk Oleksandr N. – Doctor of Technical Sciences, Professor, Professor of the Software Engineering Department, Head of the Software Engineering Department, Vinnytsia National Technical University, Vinnytsia, e–mail: rom8591@gmail.com.