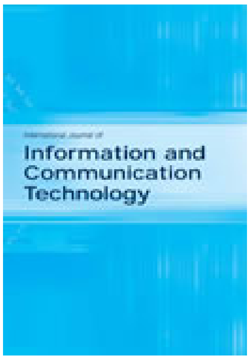


International Journal of Information and Communication Technology

ISSN online: 1741-8070 - ISSN print: 1466-6642
https://www.inderscience.com/ijict

The application of big data analysis in logistics supply chain optimisation

Chunsheng Liu

Article History:

Received:	27 September 2024
Last revised:	10 October 2024
Accepted:	11 October 2024
Published online:	13 December 2024

The application of big data analysis in logistics supply chain optimisation

Chunsheng Liu

Jiangxi Institute of Fashion Technology,
Nanchang 330201, China
Email: liuchunsheng@jift.edu.cn

Abstract: This article studies the application of big data in logistics supply chain optimisation, combining modern data processing technology with intelligent optimisation algorithms, aiming to improve supply chain efficiency and reduce logistics costs. The K-means clustering algorithm is used to partition the delivery area and customer demand, in order to optimise the allocation of logistics resources and warehouse layout, and reduce the redundancy of transportation paths. Next, based on the clustering results, the ant colony algorithm is utilised to address the optimisation of vehicle routing problem (VRP), finding the shortest path between multiple delivery points to minimise transportation time and cost. This article utilises the big data analysis platform Hadoop for data storage and processing, ensuring the efficient operation of algorithms on large-scale data. The results show that the supply chain optimisation strategy combining big data analysis, K-means clustering, and ant colony optimisation can improve delivery efficiency and reduce costs.

Keywords: big data; ant colony algorithm; ACO; K-means; logistics chain; vehicle routing problem; VRP.

Reference to this paper should be made as follows: Liu, C. (2024) 'The application of big data analysis in logistics supply chain optimisation', *Int. J. Information and Communication Technology*, Vol. 25, No. 9, pp.104–116.

Biographical notes: Chunsheng Liu received his Master's degree from the Jiangxi Agricultural University in 2024. He is currently a Lecturer at the Jiangxi Institute of Fashion Technology and his current research interest is in logistics system regulation and supply chain management.

1 Introduction

With the advancement of globalisation and the diversification of market demand, logistics supply chain plays an increasingly important role in modern enterprise operations. Supply chain management not only covers various links from raw material procurement to product distribution, but also involves multiple complex operational areas such as inventory management, transportation planning, and warehouse layout. Effective supply chain management can significantly improve a company's market responsiveness and competitiveness. However, traditional supply chain management methods often fail to fully cope with the dynamic changes, complexity, and uncertainty of today's market, especially in the face of growing customer demands, expanding logistics networks, and optimising resource allocation. Therefore, big data analysis technology provides new possibilities for solving these problems.

In recent years, the swift advancement of big data technology has led to unprecedented opportunities for optimising logistics supply chains. By analysing massive amounts of data in real-time, enterprises can obtain more accurate decision support, thereby optimising costs, improving efficiency, and enhancing flexibility in supply chain management. The data sources in the logistics supply chain are very diverse, covering order data, transportation data, warehousing data, and customer behaviour data. How to fully utilise this data for supply chain optimisation services is an important topic in current research (Wang et al., 2016). With the help of big data analysis technology, enterprises can timely understand market trends, optimise inventory management, dynamically adjust transportation plans, and better meet customer needs (Queiroz and Telles, 2018).

Big data analysis can help businesses identify potential optimisation opportunities by collecting, processing, and analysing various types of data in the supply chain. For example, based on historical order data, companies can predict future demand fluctuations and adjust production and distribution strategies based on these predictions. For example, by analysing transportation data, vehicle scheduling, delivery routes, and transportation time can be optimised, further reducing logistics costs. In addition, big data analysis technology can also help enterprises achieve visual management of different stages within the supply chain, enabling managers to monitor the operational status of the supply chain in real-time, enabling timely adjustments to strategies based on data analysis results. Li et al. (2015) optimised key operations like inventory management, logistics distribution, and demand forecasting through a big data analysis platform combined with supply chain data mining technology.

The application of intelligent optimisation algorithms in supply chain optimisation, especially in the big data environment, is particularly important. Traditional optimisation methods are often unable to handle complex dynamic changes in logistics supply chains, while intelligent optimisation algorithms such as K-means clustering algorithm and ant colony algorithm (ACO) have significant advantages in dealing with these complex problems. The K-means algorithm can effectively perform cluster analysis on customer needs, warehouse locations, and transportation networks, thereby achieving optimised resource allocation. As a heuristic algorithm based on the natural foraging behaviour of ants, ACO is particularly suitable for solving optimisation of vehicle routing problems (VRP) and can find the optimal solution for delivery routes in big data scenarios.

Specifically, the application of K-means clustering algorithm in logistics supply chain is first manifested in the division of customer demand. There are significant differences among different customer groups in terms of geographic location, order volume, purchase frequency, etc. Traditional delivery strategies often cannot effectively meet the needs of all customers. Through K-means clustering, companies can group customers with similar needs into clusters and design more accurate delivery plans. For example, based on K-means clustering results, enterprises can divide their distribution network into several regions, where customer demands are relatively consistent within each region, making it easier for centralised management and resource scheduling. In addition, K-means clustering can also be used for warehouse layout optimisation, by analysing the geographical distribution of customer groups to determine the optimal location of warehouses, in order to reduce delivery time and costs.

The ACO is employed to solve the VRP in logistics supply chain. Within each clustering area, enterprises need to plan the optimal delivery route to ensure that vehicles can cover all delivery points with the shortest path. The ACO can determine the optimal solution in complex distribution networks by mimicking the behaviour of ants as they search for the best path while foraging. Unlike traditional path planning algorithms, ACO can adapt to dynamically changing environments, such as changes in real-time traffic conditions, increase or cancellation of orders, etc. Therefore, it exhibits strong flexibility and adaptability in practical applications of logistics supply chain.

López-Santana et al. (2018) proposed a hybrid system combining K-means clustering with ant colony optimisation for path planning and scheduling problems in express delivery services, significantly improving the overall efficiency of logistics distribution. Kuo et al. (2005) explored the effectiveness of combining ACO with K-means clustering, demonstrating how to improve clustering quality by optimising initial clustering centres, especially when dealing with large-scale logistics data. Calabrò et al. (2020) proposed a path optimisation method based on ACO through simulation optimisation, which performs well in handling large-scale logistics and routing planning problems and can significantly enhance the operational efficiency of logistics systems. The ACO first proposed by Dorigo et al. (1996) has received widespread attention in recent years due to its adaptability to dynamic environments. Many studies have shown that ACO can find high-quality solutions when solving VRP. Anitha and Patil (2018) proposed an intelligent optimisation scheme for supply chain based on big data analysis, which combines clustering and optimisation algorithms to achieve intelligent management of inventory management and distribution path planning.

In the actual operation of logistics supply chain, the dynamic changes of order data and transportation data often make optimisation problems more complex (Lai et al., 2018; Nguyen et al., 2018). Therefore, this article chooses the big data analysis platform Hadoop to process massive amounts of data, ensuring the efficiency and reliability of data processing and analysis. Hadoop, as an open-source distributed computing framework, can quickly process and store large-scale data, making it particularly suitable for executing complex algorithm calculations and real-time analysis in big data environments. Through the Hadoop platform, data in the supply chain can be effectively integrated and provide support for subsequent optimisation algorithms.

This article validates the effectiveness of the suggested solution through the analysis of a logistics supply chain case. The results indicate that the supply chain optimisation strategy combining big data analysis, K-means clustering, and ant colony optimisation can significantly improve delivery efficiency, reduce transportation costs, and improve customer satisfaction. Through big data analysis technology, enterprises can obtain more accurate and real-time decision support in supply chain management, and intelligent optimisation algorithms further enhance the intelligence level of path planning and resource allocation.

This article has made the following main contributions in the field of logistics supply chain optimisation:

1 Combining K-means clustering algorithm for distribution area optimisation: this article is the first to apply K-means clustering algorithm to distribution area division and customer demand classification in logistics supply chain. By clustering analysis of customer needs, geographical location, and other factors, the warehouse layout

and resource allocation have been optimised, reducing redundant paths and resource waste in logistics transportation.

2 Introduce a path optimisation approach based on the ACO: this paper uses ACO to solve the VRP in logistics distribution. This algorithm can find the shortest path between multiple delivery points while adapting to dynamic changes in actual situations, such as changes in traffic conditions and order changes. Compared with traditional path planning methods, ACO exhibits better adaptability and optimisation effects in complex environments.

3 Application and integration of big data platforms: this article uses the Hadoop big data platform to process and analyse massive amounts of data in the supply chain, ensuring the efficient operation of algorithms in a big data environment. This platform provides excellent data processing and computing power support for the clustering and optimisation algorithm proposed in this article, ensuring the scalability and wide applicability of the scheme.

2 Relevant technologies

2.1 K-means algorithm

The K-means algorithm mainly used for data clustering analysis (Likas et al., 2003; Sinaga and Yang, 2020). The core objective of this algorithm is to divide the dataset into K clusters and, by continuously adjusting the cluster centres, assign each data point to the nearest cluster's centroid, thereby minimising the distance between data points within the cluster and the centroid (Yu et al., 2018; Zeebaree et al., 2017).

The specific steps of K-means algorithm:

1 Initialisation: randomly choose K data points as the initial centroids.

2 Assign data points. Measure the distance from each data point to the centroids and allocate each point to its closest centroid, usually using Euclidean distance to calculate the distance:

$$d\left(x_i, \mu_j\right) = \sqrt{\sum_{k=1}^{n} \left(x_{ik} - \mu_{jk}\right)^2} \tag{1}$$

where x_i represents the data point, μ_j represents the centroid, and n is the dimension of the data.

3 Update centroid. Recalculate the centroid of each cluster and update it to the average of all data points within that cluster:

$$\mu_j = \frac{1}{\left|C_j\right|} \sum_{x_i \in C_j} x_i \tag{2}$$

where C_j represents all data points in cluster j, $|C_j|$ is the number of data points in cluster j, and μ_j is the update.

4 Repetitive iteration. Continuously repeating the steps of assigning data points and updating centroids until the position of the centroids no longer changes or until the centroid reaches the maximum number of iterations.

5 Objective function. The objective of the K-means algorithm is to minimise the sum of squared distances between data points in a cluster and their centroid, and its optimisation objective function is:

$$J = \sum_{j=1}^{K} \sum_{x_i \in C_j} \left\| x_i - \mu_j \right\|^2 \tag{3}$$

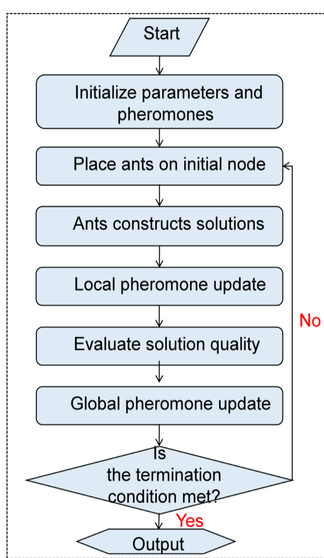
where J represents the sum of squared errors within all clusters, and the algorithm minimises this value by continuously adjusting the position of the centroid, μ_j is an update.

After multiple iterations, the algorithm usually converges to a local optimum and the centroid no longer moves.

2.2 Ant colony

The ACO is a heuristic optimisation method inspired by the natural foraging behaviour of ants, first proposed by Italian scholars in 1992. ACO is mainly used to address combinatorial optimisation problems, including the travelling salesman problem (TSP), VRP, and others (Martens et al., 2007; Pedemonte et al., 2011). In nature, ants secrete pheromones to mark the quality of pathways, and subsequently tend to choose pathways with higher concentrations of pheromones. ACO imitates this behaviour by constructing paths, updating pheromones, and performing optimisation iterations to gradually approach the optimal solution (Shelokar et al., 2004; Parpinelli et al., 2002). The basic flowchart of ACO is shown in Figure 1.

Figure 1 ACO flowchart (see online version for colours)



1 Initialisation phase. Initialise the pheromone value T_{ij} on each path and set it as an initial constant T_0:

$$T_{ij} = T_0, \forall(i, j) \tag{4}$$

In practical applications, the initial value of pheromones is usually set to a small positive value to ensure that all paths have the possibility of selection.

2 Path selection stage. Ants construct paths based on pheromones and heuristic factors. At each node, ants will select the next node based on the following probability equation:

$$P_{ij} = \frac{[T_{ij}]^{\alpha} [\eta_{ij}]^{\beta}}{\sum_{k \in allowed} [T_{ik}]^{\alpha} [\eta_{ik}]^{\beta}} \tag{5}$$

where T_{ij} represents the concentration of pheromones from node i to node j; $\eta_{ij} = \frac{1}{d_{ij}}$ is the heuristic factor from node i to j (usually the reciprocal of the distance); α and β are weight parameters that regulate the degree of influence of pheromones and heuristic factors.

3 Pheromone update stage. After constructing the path, each ant will update the pheromone of the path it has passed through. The update of pheromones includes the volatilisation of pheromones and the deposition of new pheromones.

The volatilisation of pheromones controls the decay of pheromones, simulating the process of pheromones dissipating over time in nature. The volatilisation equation is:

$$T_{ij} \leftarrow (1 - \rho) T_{ij} \tag{6}$$

where ρ is the volatility coefficient of pheromones, $0 < \rho < 1$.

Ants passing through the path add new pheromones to the path based on its quality, and the increment of pheromones is inversely proportional to the length of the path:

$$\Delta_{ij}^{k} = \frac{Q}{L_k} \tag{7}$$

where L_k is the total length of the path taken by the k^{th} ant, and Q is a constant. Q is generally set to a positive number and is usually used to adjust the strength of pheromones to affect the effectiveness of ant path selection.

After completing the pheromone update, the new pheromone concentration is:

$$T_{ij} \leftarrow (1 - \rho) T_{ij} + \Delta T_{ij}^{k} \tag{8}$$

4 Path construction and pheromone reinforcement. Each ant completes the entire path construction by gradually selecting path nodes, and the total length of the path can be calculated using the following equation:

$$L = \sum_{(i,j) \in path} d_{ij} \tag{9}$$

where d_{ij} is the distance of path $i \rightarrow j$.

To avoid the algorithm getting stuck in local optima too early, an elite strategy can be used to strengthen the pheromones on the global optimal solution path:

$$T_{ij} \leftarrow T_{ij} + \frac{Q}{L_{best}} \tag{10}$$

where L_{best} is the global optimal path length in the current iteration.

5 Termination conditions. The termination condition of ACO is usually set as: reaching the maximum number of iterations or pheromone convergence (i.e., pheromone updates tend to be stable and the optimal path no longer changes).

If the termination condition is satisfied, the algorithm outputs the optimal path; otherwise, it returns to step 2 for another mm iteration.

3 Logistics supply chain optimisation model based on K-means clustering and ACO

This method combines K-means clustering algorithm and ACO for logistics distribution area partitioning and VRP. Firstly, the delivery area is divided into several clusters using the K-means algorithm, and then the ACO is used to find the optimal delivery path within each cluster to minimise transportation time and cost (Chi and Yang, 2008; Ran et al., 2024).

The K-means algorithm is used to divide delivery points into K clusters to help optimise the allocation of logistics resources.

Randomly select K centroids (initial centroids) $\mu_1, \mu_2, \ldots, \mu_k$, calculate the Manhattan distance between each data point x_i and each centroid μ_j (different from the Euclidean distance, used to reduce the influence of outliers):

$$d\left(x_i, \mu_j\right) = \sum_{k=1}^{n} \left| x_{ik} - \mu_{jk} \right| \tag{11}$$

where x_i is the data point, μ_j is the centroid, and n is the dimension.

Update the centroid position of each cluster and use weighted averaging to adjust each centroid:

$$\mu_j = \frac{\sum_{x_i \in C_j} w_i \cdot x_i}{\sum_{x_i \in C_j} w_i} \tag{12}$$

where w_i is the weight of each point.

Iterate and repeat the above steps until the centroid no longer changes or the magnitude of the centroid change is less than the preset threshold:

$$\Delta\mu_j < \sigma \tag{13}$$

where σ is a small positive number used to determine whether the centroid converges.

Minimise the weighted sum of squares of Manhattan distances within a cluster:

$$J = \sum_{j=1}^{K} \sum_{x_i \in C_j} w_i \cdot \left| x_i - \mu_j \right|^2 \tag{14}$$

This equation can be used to optimise the division of logistics areas and reduce path redundancy.

After completing the regional division, the ACO is used to optimise the delivery paths within each cluster, with the goal of finding the optimal path and minimising the distance and time travelled by vehicles.

Initialisation of pheromone: initialise pheromone matrix T_{ij} with an initial value of constant T_0:

$$T_{ij} = T_0 \tag{15}$$

where i and j represent two delivery points.

Using roulette wheel selection method to select the next node based on pheromone concentration and distance, the selection probability P_{ij} is given by the path selection function:

$$P_{ij} = \frac{[T_{ij}] \cdot (h_{ij})^{\beta}}{\sum_{k \in allowed} [T_{ik}] \cdot (\eta_{ik})^{\beta}} \tag{16}$$

where $h_{ij} = \frac{1}{t_{ij}}$ represents heuristic information (usually the reciprocal of travel time), and β controls the degree of influence of heuristic information.

After each ant completes the path construction, local pheromone updates are performed to avoid all ants being concentrated on the same path:

$$T_{ij} = (1 - \varepsilon) T_{ij} + \varepsilon T_0 \tag{17}$$

where ε is the local pheromone update parameter, usually a small positive value.

After completing the path construction for all ants, update the global pheromone. Increase in pheromones on the optimal path:

$$T_{ij} = (1 - \rho) T_{ij} + \rho \cdot \frac{Q}{L_{best}} \tag{18}$$

where ρ is the pheromone volatilisation coefficient, Q is a constant, and L_{best} is the total distance of the optimal path.

VRPs typically also include path constraints, such as vehicle capacity or time window constraints. This can be expressed by incorporating a penalty term into the objective function:

$$f(x) = \sum_{(i,j) \in path} d_{ij} + \lambda \cdot penalty(x) \tag{19}$$

where the penalty function is used to impose additional costs on paths that violate constraints, with λ being the penalty coefficient.

The total distance of each ant's path is calculated as follows:

$$L_k = \sum_{(i,j)\in path} d_{ij} \tag{20}$$

where d_{ij} is the distance between node i and node j.

If the number of vehicles is optimised simultaneously, the goal can be extended to minimise the number of vehicles used V and the total path distance L:

$$\min(V + \alpha \cdot L) \tag{21}$$

where α is the weight that balances the number of vehicles and path length.

Iterate to construct paths and update pheromones until the stopping condition is reached, such as a maximum of $T_{\max}$ iterations or the optimal path remains stable:

$$t > T_{\max} \text{ or } \Delta L_{best} < \varepsilon \tag{22}$$

If the path length of the global optimal solution no longer changes after several iterations, the algorithm stops:

$$\Delta L_{best} = 0 \tag{23}$$

The ultimate objective function is to minimise the sum of the total path distance and the penalty for violating constraints:

$$f_{total} = \sum_{k=1}^{m} L_k + \sum_{k=1}^{m} \lambda \cdot penalty(x_k) \tag{24}$$

where L_k is the total distance of each ant's path.

Optimising the partitioning of delivery areas through K-means algorithm, combined with ACO for vehicle path optimisation in each area, can significantly reduce path redundancy, improve efficiency, and minimise total transportation costs in logistics distribution.

4 Experimental results and analysis

4.1 Data processing and experimental setup

In order to address the challenges of large-scale datasets, this study uses Apache Hadoop as the core data processing platform. In order to meet the requirements of large-scale data processing, the Apache Hadoop experimental environment in this study adopts a distributed cluster architecture, which includes a main node (NameNode) and multiple data nodes (DataNodes). The master node is responsible for managing the metadata of the file system, equipped with multi-core processors, over 16 GB of memory, and high-speed SSDs. Data nodes are used to store actual data, with each node equipped with a processor of at least 4 cores, 16 GB of memory, and at least 2 TB of storage capacity. All nodes are connected via Gigabit Ethernet to ensure efficient data transmission and processing performance. This environment supports processing over a million order records, vehicle operation logs, and warehouse data. Hadoop's distributed file system (HDFS) is used to store massive amounts of order data, transportation data, and warehousing data involved. The dataset covers over one million order records, each of which includes dimensions such as timestamp, geographic location, and order volume. The transportation data includes the operation records of 100 vehicles, with an average daily travel distance of approximately 150 kilometres per vehicle, totalling over 15,000 vehicle operation logs. Storage data involves the storage capacity and geographic location information of 50 warehouses. Using Hadoop's MapReduce framework, we performed the following data pre-processing operations:

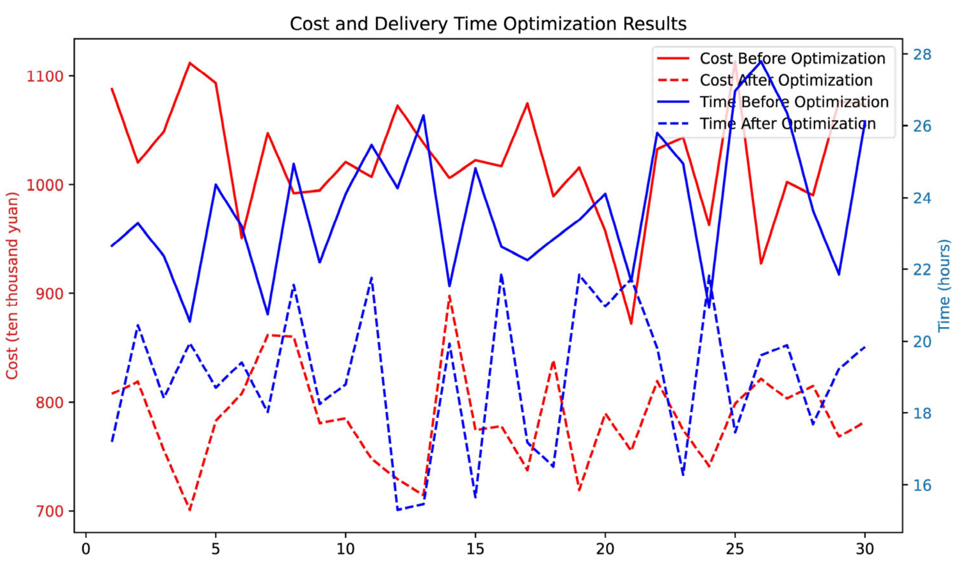
1. Data cleaning: removing duplicate and incomplete records, such as missing geographic coordinate entries in order data.
2. Data conversion: standardised geographic coordinate format and unified timestamp representation.
3. Data fusion: integrate order data and transportation data, map orders to the nearest warehouse to optimise subsequent route planning.

4.2 Experimental design

After data pre-processing, this study first uses the K-means clustering algorithm to determine the optimal distribution centre layout. Cluster analysis is based on the geographical location of orders, with an initial set of 20 cluster centres and 50 iterations to ensure convergence. Based on the clustering results, optimise the geographical layout of the warehouse and reduce the total delivery distance.

The ant colony optimisation algorithm was then applied to optimise the vehicle delivery route, with 100 ants and 200 iterations set for the process. By simulating the path discovery mechanism of ants searching for food, search for the lowest cost delivery route. The parameter settings of ACO include pheromone evaporation rate of 0.5, pheromone intensity of 0.1, and a weight of 2 for the heuristic function.

Figure 2 Experimental result chart (see online version for colours)



4.3 Analysis of experimental results

Figure 2 shows the optimisation results of cost and delivery time. The red solid and dashed lines indicate the daily transportation costs before and after optimisation, while the blue solid and dashed lines represent the delivery times before and after optimisation. From the graph, it can be seen that as the number of day's increases, the optimisation measures significantly reduce transportation costs and shorten delivery time, which proves the effectiveness of the adopted method.

Figures 3 and 4 shows the comparison results between optimisation methods and traditional methods in terms of cost and delivery time:

Figure 3 Cost comparison chart (see online version for colours)

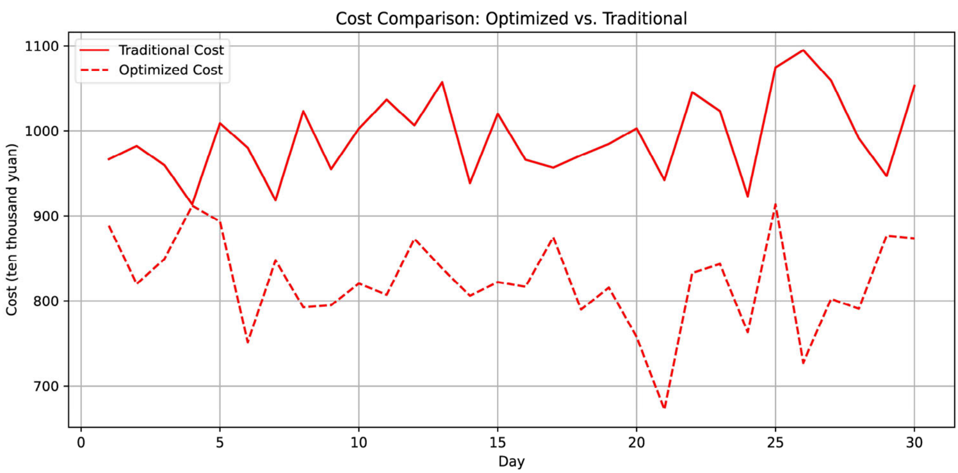


Figure 4 Delivery time comparison chart (see online version for colours)

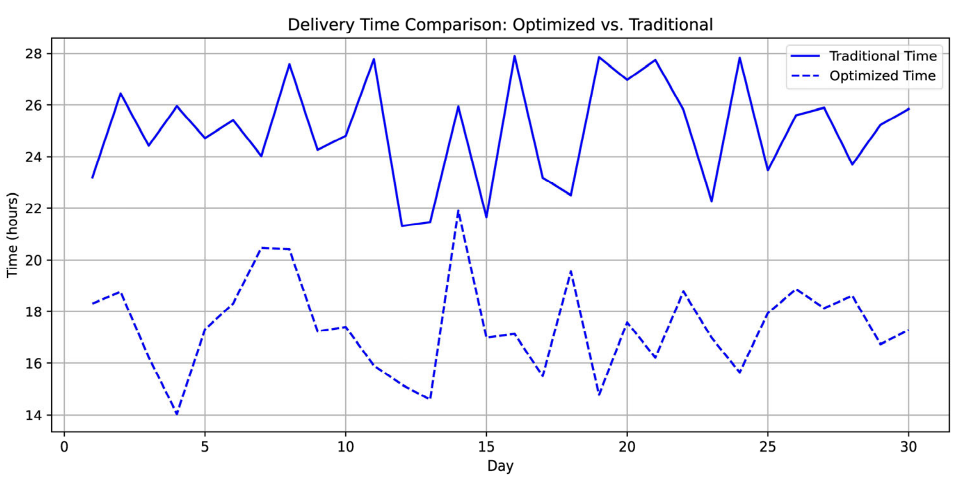


Figure 3 shows a cost comparison chart, where the red solid line represents the daily transportation cost of the traditional method, and the red dashed line represents the cost after adopting the optimisation method. It can be clearly seen that the cost of optimisation methods is generally lower than traditional methods, demonstrating significant improvements in economic efficiency.

Figure 4 shows a comparison of delivery times, with the blue solid line representing the delivery time of the traditional method and the blue dashed line representing the delivery time of the optimised method. The results also showed that the optimisation method significantly reduced delivery time and improved service efficiency. The results show that by applying big data analysis, K-means clustering, and ant colony optimisation algorithms, the method proposed in this study has significant advantages in reducing transportation costs and delivery time compared to traditional methods. On average, it reduces transportation costs by about 20% and delivery time by about 25% through optimisation, which may increase customer satisfaction and improve order turnover.

5 Conclusions

This article successfully applies big data analysis technology to optimise the logistics supply chain. By integrating modern data processing technology and intelligent optimisation algorithms, it significantly improves the operational efficiency of the supply chain and reduces logistics costs. Through detailed data analysis and algorithm application, this study not only theoretically demonstrates the potential of big data technology in supply chain management, but also verifies the practical value of these technologies through practical case analysis. Firstly, this article effectively processes massive data in the supply chain using the Hadoop platform, including order data, transportation data, and warehousing data, ensuring the efficiency and scalability of the processing process. Secondly, by implementing the K-means clustering algorithm, the layout of distribution areas and warehouses has been optimised, which not only reduces redundancy on transportation routes but also improves the rationality of resource allocation. Furthermore, this article uses ACO to optimise the delivery route of vehicles, effectively solving the VRP and achieving the shortest path search between multiple delivery points, thereby minimising transportation time and cost. The experimental results clearly indicate that the supply chain management strategy proposed in this study, which combines big data analysis, K-means clustering, and ant colony optimisation, can effectively improve delivery efficiency and reduce operating costs. These achievements not only have direct guiding significance for the practical operation of the logistics industry, but also provide new perspectives and technological approaches for future supply chain management research. Subsequent research can enhance the dynamic response capability and optimisation effect of supply chain management by integrating real-time data and introducing hybrid optimisation algorithms. Meanwhile, attention should also be paid to the application of green logistics, sustainable development, and intelligent prediction technology to further reduce costs and enhance the environmental friendliness and efficiency of the supply chain.

Acknowledgements

This work is supported by the Research Center of Digital Logistics and apparel supply chain: Research and Innovation Team of Intelligent Logistics.

References

Anitha, P. and Patil, M.M. (2018) 'A review on data analytics for supply chain management: a case study', *International Journal of Information Engineering and Electronic Business*, Vol. 10, No. 5, p.30.

Calabrò, G., Torrisi, V., Inturri, G. et al. (2020) 'Improving inbound logistic planning for large-scale real-world routing problems: a novel ant-colony simulation-based optimization', *European Transport Research Review*, Vol. 12, pp.1–11.

Chi, S-C. and Yang, C-C. (2008) 'A two-stage clustering method combining ant colony SOM and K-means', *Journal of Information Science & Engineering*, Vol. 24, p.5.

Dorigo, M., Maniezzo, V. and Colorni, A. (1996) 'Ant system: optimization by a colony of cooperating agents', *IEEE transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, Vol. 26, No. 1, pp.29–41.

Kuo, R., Wang, H., Hu, T-L. et al. (2005) 'Application of ant K-means on clustering analysis', *Computers & Mathematics with Applications*, Vol. 50, Nos. 10–12, pp.1709–1724.

Lai, Y., Sun, H. and Ren, J. (2018) 'Understanding the determinants of big data analytics (BDA) adoption in logistics and supply chain management: an empirical investigation', *The International Journal of Logistics Management*, Vol. 29, No. 2, pp.676–703.

Li, J., Tao, F., Cheng, Y. et al. (2015) 'Big data in product lifecycle management', *The International Journal of Advanced Manufacturing Technology*, Vol. 81, pp.667–684.

Likas, A., Vlassis, N. and Verbeek, J.J. (2003) 'The global k-means clustering algorithm', *Pattern Recognition*, Vol. 36, No. 2, pp.451–461.

López-Santana, E., Rodríguez-Vásquez, W. and Méndez-Giraldo, G. (2018) 'A hybrid expert system, clustering and ant colony optimization approach for scheduling and routing problem in courier services', *International Journal of Industrial Engineering Computations*, Vol. 9, No. 3, pp.369–396.

Martens, D., De Backer, M., Haesen, R. et al. (2007) 'Classification with ant colony optimization', *IEEE Transactions on Evolutionary Computation*, Vol. 11, No. 5, pp.651–665.

Nguyen, T., Li, Z., Spiegler, V. et al. (2018) 'Big data analytics in supply chain management: a state-of-the-art literature review', *Computers & Operations Research*, Vol. 98, pp.254–264.

Parpinelli, R.S., Lopes, H.S. and Freitas, A.A. (2002) 'Data mining with an ant colony optimization algorithm', *IEEE Transactions on Evolutionary Computation*, Vol. 6, No. 4, pp.321–332.

Pedemonte, M., Nesmachnow, S. and Cancela, H. (2011) 'A survey on parallel ant colony optimization', *Applied Soft Computing*, Vol. 11, No. 8, pp.5181–5197.

Queiroz, M.M. and Telles, R. (2018) 'Big data analytics in supply chain and logistics: an empirical approach', *The International Journal of Logistics Management*, Vol. 29, No. 2, pp.767–783.

Ran, X., Suyaroj, N., Tepsan, W. et al. (2024) 'A hybrid genetic-fuzzy ant colony optimization algorithm for automatic K-means clustering in urban global positioning system', *Engineering Applications of Artificial Intelligence*, Vol. 137, p.109237.

Shelokar, P., Jayaraman, V.K. and Kulkarni, B.D. (2004) 'An ant colony approach for clustering', *Analytica Chimica Acta*, Vol. 509, No. 2, pp.187–195.

Sinaga, K.P. and Yang, M-S. (2020) 'Unsupervised K-means clustering algorithm', *IEEE Access*, Vol. 8, pp.80716–80727.

Wang, G., Gunasekaran, A., Ngai, E.W. et al. (2016) 'Big data analytics in logistics and supply chain management: certain investigations for research and applications', *International Journal of Production Economics*, Vol. 176, pp.98–110.

Yu, S-S., Chu, S-W., Wang, C-M. et al. (2018) 'Two improved k-means algorithms', *Applied Soft Computing*, Vol. 68, pp.747–755.

Zeebaree, D.Q., Haron, H., Abdulazeez, A.M. et al. (2017) 'Combination of K-means clustering with genetic algorithm: a review', *International Journal of Applied Engineering Research*, Vol. 12, No. 24, pp.14238–14245.