https://doi.org/10.31449/inf.v48i14.6030 Informatica **48** (2024) 171–188 **171**

Travel Path Recommendation Algorithm Based on Hybrid Particle Swarm and Ant Colony for Social Media Shared Data Mining

Kun Zhu
School of Culture and Tourism, Jilin Province Economics and Management Cadres College, Changchun 130000, China
E-mail: zk1234561206@163.com

Keywords: PSO, ant colony recommended tourist routes, social media, recommended accuracy

Received: April 16, 2024

The effective mining of shared data on social media can help personalized recommendations of tourist attractions and paths for users. This study proposes a tourism path recommendation scheme that combines PSO and ant colony optimization to address the issue of low recommendation accuracy caused by incomplete extraction of effective information in tourism path recommendation algorithms. The tourism path recommendation algorithm obtains a pseudo demand sequence based on the distance between the user's center point, and obtains attribute keywords through the user's evaluation text and text extraction technology. The algorithm employs the iterative operation of the particle swarm ant colony algorithm to determine the semantic distance and geographic distance of the target user to the optimal sequence, and updates the preference distance through a weighted calculation. For the four benchmark functions, the proposed algorithm had a longer running time under the same number of runs. Under the four benchmark test functions f1, f2, f3, and f4, when the maximum number of runs was reached, the running time of the algorithm was 36.58s, 62.96s, 90.59s, and 64.26s, respectively. The proposed PSO-AC travel path recommendation algorithm had lower recommendation errors under different running times, and the range of error values for route recommendation was 0.005-0.089. In the training set, the confusion matrix results of the algorithm showed that the accuracy of tourism path recommendation for topics 1 and 5 was 81.25% and 84.26%, respectively, and the recommendation accuracy for the other three topics was also above 75%. The designed algorithm takes into account both emotional and time series dimensions, and has high recommendation accuracy. It has obvious advantages in the actual process of recommending tourist routes.

Povzetek: Predlagan je nov algoritem priporočanja turističnih poti, ki združuje optimizacijo rojev delcev (PSO) in optimizacijo kolonije mravelj za rudarjenje podatkov z družbenih omrežij.

1 Introduction

Social media, as an important communication medium in the era of new media, has significant data mining value in various fields such as traffic management, urban planning, and traffic prediction. Social media platforms cover various multi-modal data of user posts and evaluations. How to extract word feature information from text information for data mining is currently a hot research topic. At the same time, there are situations in social media platforms where some users do not have or only have partial text information, which makes it difficult to obtain valuable text information from historical information, resulting in many text analysis tasks being unable to be completed smoothly. The application of data mining and the processing of social media network information, including the geographical location and ticket prices of relevant attractions, enables the provision of constructive guidance for personalized tourism based on user preferences. This, in turn, facilitates the realization of the personalized needs of tourists [1-3]. The recommendation of tourist routes is an important service in tourism service information. According to data from tourism service websites, some tourism agencies offer pre-booked travel routes without considering the personalized needs of users. It is challenging for users to swiftly identify a travel route that aligns with their specific preferences amidst the vast array of tourism information. This process not only consumes a considerable amount of personal time but also results in a sub-optimal travel experience for users. A substantial corpus of domestic and foreign literature exists on the topic of personalized recommendations for tourist attractions and paths. However, research results have not yet fully explored the potential of user social media information, which has resulted in unsatisfactory recommendation outcomes. Moreover, there is no established algorithm model that combines user semantic text information to analyze recommendation schemes [4-6]. Particle swarm optimization (PSO) has the characteristic of high recommendation efficiency in path recommendation algorithm (PRA), and ant colony (AC) algorithm can compensate for the shortcomings of PSO. In response to the shortcomings of existing personalized recommendation algorithms in user preference information data mining and the problems in recommendation scheme analysis, this study proposes a travel path recommendation algorithm (TPRA) that combines PSO and AC algorithm, namely the PSO-AC algorithm. The research aims to provide feasible solutions for fully mining effective information in social media shared data (SMSD). The research content is elaborated through four parts. Part 1 is an analysis of the present application status of SMSD, PSO, and AC in personalized recommendations. Part 2 has designed a recommendation plan for tourism routes based on SMSD, with a focus on introducing TPRA that combines PSO and AC algorithms. Part 3 analyzes the performance of TPRA for PSO and AC, and sets up comparative recommendation algorithms to verify the effectiveness of the proposed method. Part 4 summarizes the research results and proposes the development direction for the next step of research, providing suggestions for the proposal of practical travel path recommendation (TPR) schemes.

SMSD mining and TPR combined with intelligent algorithms have been widely reported by experts. Jamal et al. came up with a TPR method based on time extended networks, which achieves personalized recommendations based on the current route congestion situation and the conditional probability value of transitioning from the current node to the next node. It transformed the route problem into a nonlinear discrete optimization problem and improved it with the idea of dynamic programming, verifying that this method had a relatively accurate recommendation effect [7]. Huang and other scholars constructed empirical and real-time traffic models to expand possible paths from the current node and calculate the cost of these candidate nodes to their destination. Based on the data calculated by the model, combined with the A* algorithm, the optimal path could be selected [8]. Javed Awan et al. observed that there is a certain degree of consistency between users' access preferences for a certain location and the frequency of visits to other locations, and proposed a probabilistic category recommendation algorithm. Compared to common PRAs, this model had higher accuracy and recall [9]. Beed et al. analyzed the optimization problem of urban tourism routes using artificial bee colony optimization algorithm and PSO. This method aimed to maximize the tourist attractions, minimize travel costs and the expected browsing time deviation. The performance of this method had been verified through standard basis functions [10]. Qamar et al. solved the traveling salesman problem using optimal and optimal ant systems and PSO, and measured the arrangement quality and assembly time numerically. The test set results showed that compared to conventional AC, this method had stronger feasibility [11].

Luo and other scholars have designed a personalized intelligent TPR technology for improved discrete PSO applications. This method analyzed tourism recommendations based on the possible tourism characteristics of different tourists, and constructed a tourist interest model based on user preferences. The recommendation accuracy of this method was about 76%, and the recommendation time was within 0.8 seconds. Therefore, it could achieve real-time accurate recommendation [12]. González et al. proposed a strategy of dynamically aggregating user preference features based on the dynamic characteristics of search paths to achieve the goals of maximizing group satisfaction and minimizing individual satisfaction differences, and based on this, established a model. However, this model relied heavily on training datasets and had over-fitting issues [13]. Zhang et al. proposed that in TPR, user preference modeling and the density of passenger distribution in scenic areas were the primary topics of concern. This often-involved user satisfaction and travel efficiency, and users often chose popular points of interest (PoI) when traveling [14]. Zhou explored the potential interests and preferences of tourists by collecting photos from social networking sites, trained convolutional neural networks, and used fuzzy set theory to generate classifiers to classify users. This method had higher recommendation accuracy compared to other recommendation algorithms [15]. Kumar et al. proposed a selective travel strategy based on user needs. It discretized the spatial structure of scenic spots, optimized PSO using AC thinking, improved the rules for location updates, and quickly found the shortest path that met the personalized needs of users and included as many scenic spots as possible [16].

There are many research reports on the analysis of time label sequences in TPR, but there are few studies on the reconstruction of interest point sequences using AC. However, due to the limited flexibility of TPR in analyzing user social relationships, few scholars have been able to fully explore the textual information of users. This study explores user-related semantic preferences through existing text information, and constructs relevant models by utilizing AC, which has advantages in combinatorial planning problems, to obtain TPRs and provide technical support for personalized TPR for tourists.

Table 1 shows the TPR methods and results of different researchers, and each recommendation method has achieved certain results in TPR. However, few researchers have considered the impact of time label sequences and interest point sequences on path recommendation. Meanwhile, due to the limited flexibility of TPRs in analyzing user social relationships, few scholars have been able to fully explore user textual information. In response to this issue, this study utilizes existing text information to mine user-related semantic preferences, and constructs relevant models using AC, which has advantages in combinatorial planning problems, to obtain TPR and provide technical support for personalized TPR for tourists.

Table 1: Summary table of related work

Author	Method	Result
Jamal et al. [7]	A tourism route recommendation method based on time expansion network	Recommendation accuracy greater than 70%
Huang et al. [8]	Build empirical traffic models and real-time traffic models, and use A * algorithm to select the best path	The recommendation accuracy is as high as 77.35%, and the recommendation time is 0.69 seconds.
Javed Awan et al. [9]	Probability category recommendation algorithm	The accuracy and recall rate of recommendations can reach over 70%.
Beed et al. [10]	Artificial bee colony optimization algorithm and PSO algorithm	The recommended accuracy is as high as 81.51%.
Qamar et al. [11]	Optimal PSO and optimal ant system	The recommendation error rate is less than 5%.
Luo et al. [12]	Improving personalized travel route intelligent recommendation technology based on discrete PSO	The recommendation accuracy is around 76%, and the recommendation time is within 0.8 seconds.
González et al. [13]	Dynamic aggregation of user preference features	There is an issue of over-fitting.
Zhang et al. [14]	User preference modeling	High recommendation accuracy.
Zhou et al. [15]	Convolutional neural networks and fuzzy set theory	The recommended accuracy is 80.65%.
Kumar et al. [16]	PSO algorithm combined with AC algorithm	The recommended time is 0.623 seconds, with a recommendation accuracy of over 80%.

2 Construction of TPR models for PSO and AC in SMSD mining

When tourists choose to travel to a certain city, they usually determine the tourist attractions through social media contribution data, and at the same time, develop travel paths based on their own preferences and needs. The mining of effective information in SMSD is crucial for personalized recommendations of tourist attractions and paths. In dealing with combinatorial programming problems, AC has significant advantages, but this algorithm is prone to falling into the phenomenon of fast convergence speed or local optima. This study proposes a TPRA, i.e., PSO-AC, which combines PSO and AC. Based on AC, this method improves the convergence speed through the maximization and minimization mechanism (Max/Min-M) and the elite ant management mechanism (EAMM). At the same time, the bidirectional encoder representation from transformers (BERT) model is used to analyze the sentiment of tourist evaluation texts, and the extracted text information is integrated into TPR.

2.1 TPR and SMSD mining

Based on a clear understanding of the set of recommended PoI in social media platforms, this study designs a travel path that meets personalized recommendations. This study first obtains the semantic vector of the sequence from the user's comment text and the semantic distance of the user to the PoI sequence. At the same time, geographic information is used to calculate geographic distance, and the preference distance is obtained through weighted processing of the two, which can be used to evaluate path quality. Subsequently, this study optimizes AC using the Max/Min-M and the EAMM, while combining them with PSO to perform combinatorial planning on the PoI recommendation point set. Tourist attractions can be regarded as PoI sequences with chronological order. TPR is considered to be the process of assembling elements from a collection of PoIs in a certain order to form a path that meets the personalized needs of users. The path recommendation problem can be quantified by selecting a point as the starting point in a weighted undirected graph. By traversing all nodes through a single access method, multiple paths are generated, and the path with the minimum total distance is given based on the preferences of the target user, which is the recommended path. The core of TPRA is to obtain a sufficient number of interest points within the search range based on the check-in data of the target user, and to obtain a quasi-demand sequence based on the distance from the user center point. At the same time, attribute keywords are obtained through user evaluation text and BERT technology. Then, the iterative operation of PSO-AC is used to obtain the semantic distance and geographic distance of the target user to the current optimal sequence, and the preference distance is updated through weighted calculation. The optimal path is the one with the shortest preferred distance, in accordance with the iteration termination conditions. Fig. 1 shows the PRA structure.

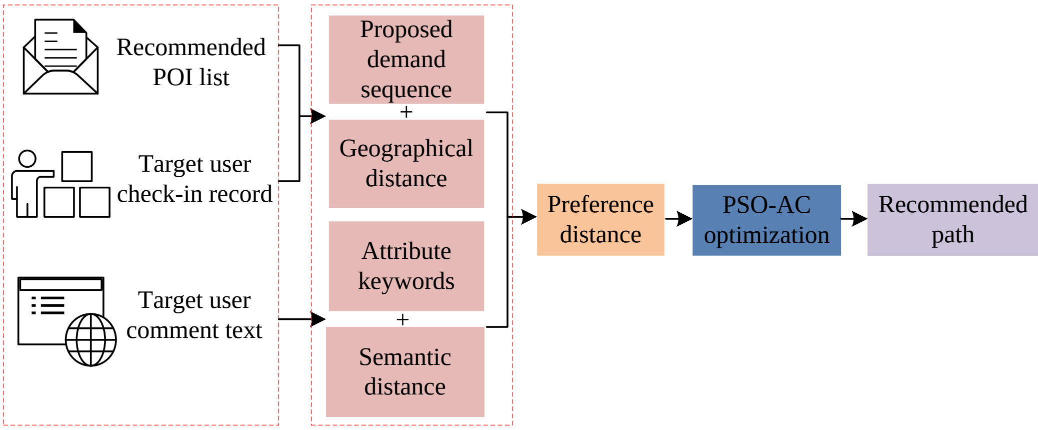


Figure 1: The structure of path recommendation algorithm

The PRA needs to combine definitions, including the distance from a point to a trajectory (D1), the user's set of quasi requirements (QRUS), sequence semantic vectors (SSV), the user's preferred geographic distance for PoI sequences (D2), the user's preferred semantic distance for PoI sequences (D3), and the user's preferred distance for PoI sequences (D4). For D1, the quantity of elements in the PoI set is set to N, which the elements can form an ordered sequence is $N!$. The distance $d_M(o,\gamma)$ between the center point of the target user and the trajectory is calculated using equation (1).

$$d_M(o,\gamma)=\min_{v_i\in\gamma}\{sd(o,v_i)\} \tag{1}$$

In equation (1), the distance between point o and point v_i is $sd(o,v_i)$, and one vertex of the trajectory is v_i. The definition of QRUS is as follows: The PoI accessed by the target user u_i forms a circular area, and QRUS is $O_{require}=\{o_1,\cdots,o_n\}$. To evaluate the correlation between PoI and label w, the PoI document is defined as I_p, with a PoI frequency of p_f, an access frequency of v_f, and a user frequency of u_f. The calculation formula for the score $AT(w)$ of a certain PoI poi_p on the label is equation (2).

$$AT(w)\propto \max_{p\in P}\frac{p_f(I_p,w)\cdot u_f(I_p,w)}{r_f(I_p,w)} \tag{2}$$

In equation (2), the threshold is $r_f(I_p,w)$. If the value exceeds the threshold $r_f(I_p,w)$, it can be considered that the label is an attribute keyword. The proposed demand for SSV is represented by $K_{require}=(K_{o_1},\cdots,K_{o_n})$, and the semantic vector of the trajectory is $K_\gamma=(K_{v_1},\cdots,K_{v_n})$. The semantic value K_{poi_p} of a certain PoI poi_p is equation (3).

$$K_{poi_p}=\frac{\sum_{i=1}^{x}AT(w_i)}{x} \tag{3}$$

The expression of D2 S_{dist} is equation (4).

$$S_{dist}=\frac{2}{1+e^{-\sum_{i=1}^{n}d_M(o,\gamma)}}-1 \tag{4}$$

This study uses the Gerald coefficient to calculate D3. Fig. 2 is a schematic diagram of the geographical distance between the target user and the PoI sequence.

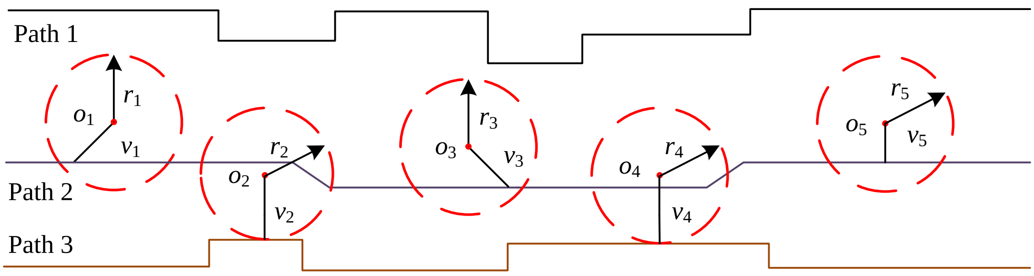


Figure 2: Geographical distance diagram between the target user and the sequence of PoI

The expression for D3 referring to $T_{dist}(K_{require},K_\gamma)$ is equation (5).

$$T_{dist}(K_{require},K_\gamma)=\frac{K_{require}\cdot K_\gamma}{\|K_{require}\|^2+\|K_\gamma\|^2-K_{require}\cdot K_\gamma} \tag{5}$$

The calculation formula for D4 is equation (6).

$$ST_{dist}\left(require,\gamma\right)= \lambda.S_{dist}\left(o_{require,}\gamma\right)+\left(1-\lambda\right).T_{dist}\left(K_{require},K_{\gamma}\right) \quad (6)$$

In equation (6), the preference related parameter is γ.

2.2 AC and elite ant management improvement mechanism

AC is a probabilistic algorithm that seeks the optimal path, with characteristics such as heuristic search, positive feedback of information, and distribution calculation. Essentially, it can be considered a heuristic global optimization algorithm [17-19]. The core idea of this algorithm is to refer to the feasible solution of the problem that needs to be optimized through the walking path of ants. All AC paths are the solution space of the matter to be optimized. In the initial stage, ants in areas with short paths release more pheromones, while over time, the number of ants choosing this path increases. Under positive feedback, the AC will all concentrate on the optimal path, and the corresponding solution is the optimal solution [20-22]. The conversion probability $p_{ij}^{k}\left(t\right)$ of ants at time t at two points i and j is expressed as equation (7).

$$p_{ij}^{k}\left(t\right)= \begin{cases} \dfrac{\gamma_{ij}^{\alpha}\left(t\right)\cdot\gamma_{ij}^{\beta}\left(t\right)}{\sum s\in allowed_{k}\gamma_{ij}^{\alpha}\left(t\right)\cdot\gamma_{ij}^{\beta}\left(t\right)} & j\in allowed_{k} \\ 0 & otherwise \end{cases} \quad (7)$$

In equation (7), the next node that the ant can choose is $allowed_{k}$. s refers to a certain node. The number of pheromone residues on the edge is γ_{ij}. a and β are information heuristic factors and expected heuristic factors, respectively. The update formula for pheromones is equation (8).

$$\gamma_{ij}\left(t+n\right)=\left(1-\rho\right)\gamma_{ij}\left(t\right)+\gamma_{ij}\left(t,t+n\right) \quad (8)$$

In equation (8), the volatilization coefficient of pheromones is ρ, and the residual coefficient of pheromones is $1-\rho$. For the updating of pheromones, this study adopts the Ant-Cycle model, which uses global information to complete pheromone updates. Compared to Ant-Density and Ant-Quantity, Ant-Cycle has significant advantages in analyzing global path planning problems. The calculation formula for Ant-Cycle is equation (9).

$$\Delta\gamma_{ij}\left(t+n\right)= \begin{cases} \dfrac{Q}{L_{k}} & if\ ant_{k}\,pass\left(i,j\right) \\ 0 & otherwise \end{cases} \quad (9)$$

In equation (10), the pheromone increment of edge $\left(i,j\right)$ is $\Delta\gamma_{ij}\left(t+n\right)$. The length of the path taken by the k-th ant is L_{k}. The intensity of pheromones is Q. Fig. 3 shows the ant optimization path optimization process.

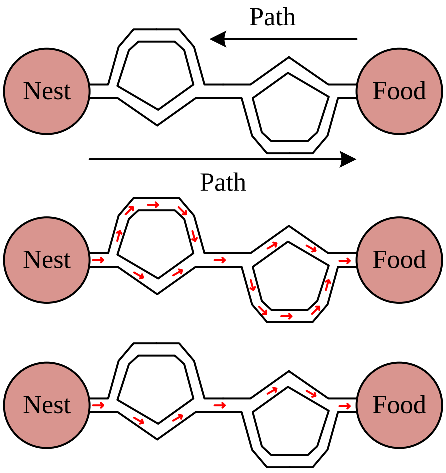


Figure 3: Ant optimization path optimization process

Due to the tendency of the algorithm to fall into the phenomenon of too fast convergence speed or local optima, this study improves the speed through the Max/Min-M and the EAMM. EAMM has been proven by scholars to increase the probability of obtaining the global optimal path while reducing convergence time. This mechanism can appropriately increase the number of additional pheromones to convert local optimal solutions into global optimal solutions. The principle for updating pheromones left over during ant foraging is equation (10).

$$\gamma_{ij}\left(t+1\right)=\rho\gamma_{ij}\left(t\right)+\Delta\gamma_{ij}+\Delta\gamma_{ij}^{*} \quad (10)$$

In equation (10), the total amount of pheromone increment left by each ant passing through edge $\left(i,j\right)$ is $\Delta\gamma_{ij}$. The formula for $\Delta\gamma^{*}$ is equation (11).

$$\square\gamma^* = \begin{cases} \frac{Q}{L_k}\sigma \; if \;\; edge(i,j) \, is \, best \, solution \\ 0 \qquad otherwise \end{cases} \tag{11}$$

In equation (11), the number of elite ants is σ, and the pheromone increment that is better than the edge (i,j) sought by the elite ants is $\square\gamma^*$. Max/Min-M only updates the pheromones of the optimal ant during each iteration, and in this case, it can search for more solutions without the phenomenon of precocity [23-25]. Meanwhile, the mechanism sets pheromone thresholds on each path to balance the differences in pheromones across different paths. When initializing pheromone values, this mechanism can search for more possible excellent solutions.

2.3 TPR Solution of PSO-AC Joint Algorithm

On the basis of AC optimization, this study analyzes the solution of tourism path combination through PSO. PSO owns lots of benefits like heuristic search, strong robustness, distributed computing, which is often applied to seek the optimal route. Fig. 4 shows the principle of AC.

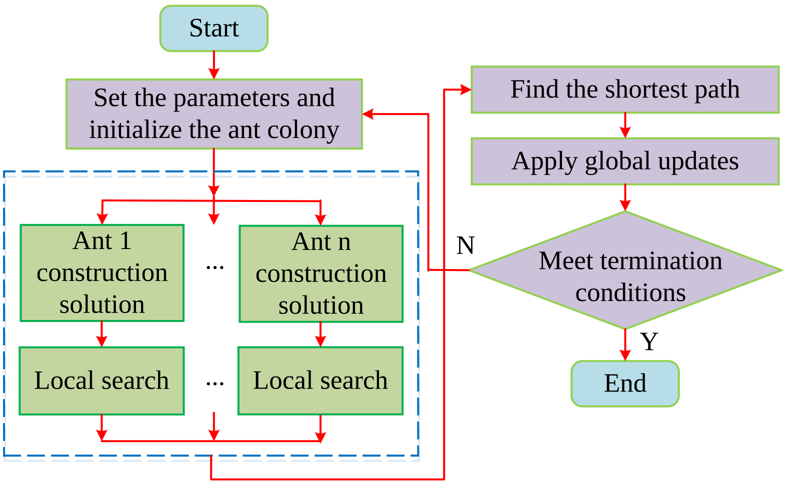


Figure 4: The principle of AC optimization algorithm

AC mimics the ants' capacity to find the shortest foraging way through information exchange. Ants secrete pheromones while foraging, which allows for complete information exchange between ant populations. The ranking order and weight of the ants increase as the path length decreases. Therefore, it is necessary to update the pheromone of the top w ants using the equation (12).

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij} + \sum_{k=2}^{w}\Delta\tau_{ij}^{k}(t) + \Delta\tau_{ij}^{*}(t) \tag{12}$$

In equation (12), the initial pheromone volatilization factor is ρ within the interval of (0,1). The second to w-th ranked ants update their pheromones as $\sum_{k=2}^{w}\Delta\tau_{ij}^{k}(t)$, and the optimal ant updates their pheromones as $\Delta\tau_{ij}^{*}(t)$. PSO can solve many linear and nonlinear matters with excellent convergence velocity, but basic PSO is more likely to getting stuck in local optima during particle search, causing the limited particle diversity. To address it, an algorithm improvement built on PSO is suggested, which enhances the particle population diversity and avoids premature stagnation and local optima [26-28]. A population of n_2 particles in the M-dimensional target search space is randomly generated. V represents the particle velocity, and U is the particle position in the search space. The corresponding fitness can be calculated. $P_i = (P_{i1}, P_{i2}, ..., P_{iM}), P_g = (P_{g1}, P_{g2}, ..., P_{gM})$ are individual and group extremums. During each iteration, particles update their velocity and location by comparing the fitness of new particles with that of the current individual and population extremum [29-30]. The velocity of the particle is determined by its current position and the particle position from the previous iteration. Equation (13) provides the update equation for the velocity.

$$\begin{aligned} V_{im}^{k+1} &= \\ & wV_{im}^{k} + c_1 r_1 \left(P_{im}^{k} - \left(1+\beta_1 U_{im}^{k-1}\right) + \beta_1 U_{im}^{k-1}\right) \\ & + c_2 r_2 \left(P_{gm}^{k} - (1+\beta_2) U_{im}^{k} + \beta_2 U_{im}^{k-1}\right) \end{aligned} \tag{13}$$

In equation (13), $m = 1,2,...M, i = 1,2,...,n_2, k$ represents the current iterations. $c_1, c_2 > 0$ is the acceleration factor. r_1, r_2 is a random number within 0 to 1. w is the inertia weight coefficient. As the iteration k increases, the nonlinearity lows down, and its expression is equation (14).

$$w = w_{ini} - (w_{ini} - w_{end})\left(k \big/ k_{\max} \right)^2 \tag{14}$$

In equation (14), w_{ini} and w_{end} are the beginning and ending values of w. Taking $\beta_i < \frac{2\sqrt{c_i - 1}}{c_i}, i = 1, 2,$ in the early phase and $\beta_i \geq \frac{2\sqrt{c_i - 1}}{c_i}, i = 1, 2,$ in the later stage of the algorithm, by enhancing global search and optimization abilities, meticulous search is achieved. The principle for updating particle positions is equation (15).

$$U_{im}^{k+1} = \begin{cases} U_{im}^{k} + V_{im}^{k+1} + 2d^k, & \sum_{i \neq j, j \neq 1}^{n} \left\| u_{im}^{k} - u_{jm}^{k} \right\| < d^k \\ U_{im}^{k} + V_{im}^{k+1}, & \sum_{i \neq j, j \neq 1}^{n} \left\| u_{im}^{k} - u_{jm}^{k} \right\| \geq d^k \end{cases} \tag{15}$$

In equation (15), $d^k = d_{ini} - (d_{ini} - d_{end})\left(\frac{k}{k_{max}}\right)^2$ is the min distance allowed between particles. d_{ini}, d_{end} and d^k are the initial and final values. Fig. 5 is a schematic diagram of PSO-AC. Firstly, the center point of the target user and the pseudo demand sequence are calculated, and based on document I_p, the attribute keywords of a PoI are calculated to obtain the semantic distance of the generated path after each iteration. Then, using the iterative operation of PSO-AC, the semantic distance and geographic distance of the target user to the current optimal sequence are obtained, and the preference distance is updated through weighted calculation. Finally, when the iteration terminates, the path with the shortest preferred distance is set as the optimal path. In the optimized ant algorithm, intelligent ants will interfere with the degree of influence during the optimization process based on the current heuristic or information factors. When the information factor has a significant influence factor, ants are highly likely to stay at the optimal path, leading to the algorithm dropping into the local optimum. When the heuristic factor has a significant impact, the algorithm will stop iterating.

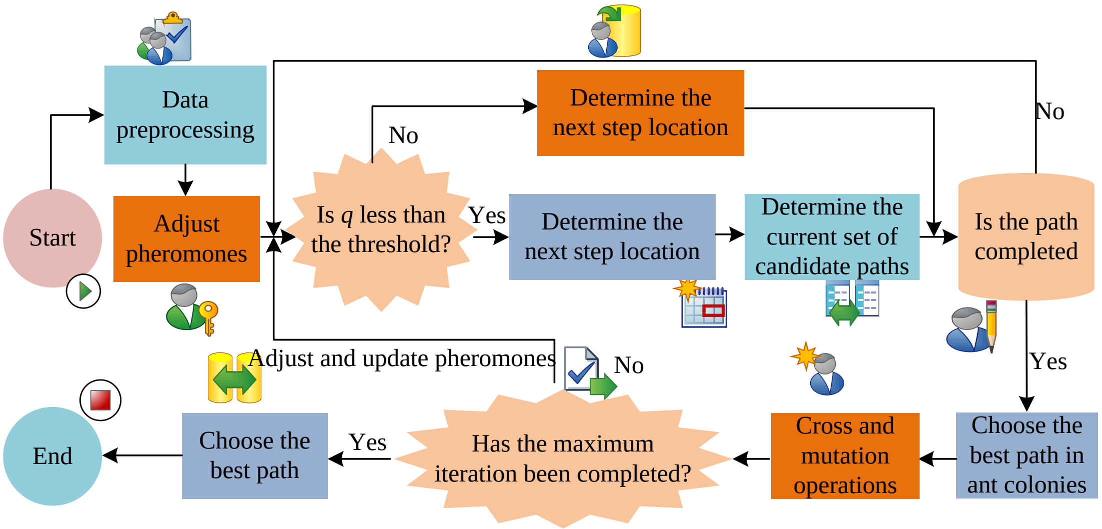


Figure 5: Schematic diagram of PSO-AC tourism PRA

3 TPRA performance analysis of PSO and AC in social media data

This study analyzed the performance of TPRA for PSO and AC in social media data, including the performance of benchmark test functions and the optimization effect of PSO-AC. The testing indicators included accuracy, running time, error value, fitness value, recommendation error, and confusion matrix.

3.1 Benchmarking the performance of benchmark functions

Experimental environment settings: The system is Windows 10, with processor of Inter(R) Core (TM) i7-6700 and 16.00GB memory. The APP version is MATLAB R2013B, and the hard disk size is 500GB. During the parameter setting process, increasing the number of particles and the maximum number of iterations can improve the search accuracy of the algorithm, but it will also increase the computational cost. Based on the complexity of the TPR problem and the limitations of computing resources, the number of particles is set to 100 and the maximum number of iterations is 400. The learning factors c1 and c2 are used to regulate the degree of influence of individual optimal solutions and global optimal solutions, respectively, affecting the velocity update process of particles. Considering the characteristics of the problem and the convergence speed of the algorithm, the learning factor is set to c1=c2=1.2. In addition, larger inertia weights make particles more inclined to move in the current direction, while smaller inertia weights make particles more inclined to be influenced by individuals and groups to change their direction of movement. Taking all factors into consideration, the maximum and minimum values of inertia weight in the study are 0.7 and 0.3, respectively. The study compares the computational complexity of PSO algorithm, AC algorithm, and TPRA of PSO-AC. The computational complexity of the PSO algorithm is O (N\cdot T), where N represents the spatial dimension of the problem and T represents the number of iterations. The computational complexity of AC algorithm is O (V2\cdot T), where V represents the number of nodes. The computational complexity of the PSO-AC TPRA is a combination of both, namely O (N\cdot T+V2\cdot T). When the number of nodes V is relatively small, PSO-AC algorithm is more efficient than using AC algorithm alone or PSO algorithm. This algorithm combines the characteristics of two algorithms and can better adapt to the complexity of the problem. The study area covers 16 administrative divisions in Beijing, and the social media data comes from WeChat, Tiktok, Xiaohongshu and other social media from 2020 to 2022. The study extracts geographic information, publication time, publication location, and text information from the data. At the same time, the text is classified into positive emotions, negative emotions, and neutral emotions through the emotion classification of text information. Subsequently, the text information is subject classified to identify content related to catering, accommodation, leisure, browsing, and weather. Tourists include both males and females. Taking months, weeks, and years as examples for statistical time. Table 2 shows the corresponding topics for different research topics.

Table 2: Corresponding topics for different research topics

Theme	Related topics	Tourist	Data sources	Data volume/10000 pieces
Catering	"Fried Liver", "Happy Tea", "Hot Pot" Restaurant/Food "Expensive", "poor service quality", service quality "Spicy", "Bad", "hungry", "unable to eat", "food allergy"	Male, Female	Sina Weibo, The Xiaohongshu, Tiktok, etc.	4.56
Stay	"Internet", "Hot Room", "Cold Room", "Poor Catering" Facilities "Poor service"	Male, Female	Sina Weibo, The Xiaohongshu, Tiktok, etc.	5.12
Leisure	"Airport Shopping" and "Beijing Marathon" "Watching shows, playing games, shopping, National Theatre, drinking coffee, and Strawberry Music Festival"	Male, Female	Sina Weibo, The Xiaohongshu	6.18
Browse	Flag raising, mountain climbing, ticket buying, check-in, and queuing activities "Cold", "Hot", "Snow", "Rain", "Wind", "Haze"	Male, Female	Tiktok	5.19
Weather	Snow, strong wind, sun, bad weather, flag raising, blue sky, sun exposure, lake freezing, sandstorms	Male, Female	Others	5.64

Fig. 6 shows the statistical results of research data for different months. The peak tourist season in Beijing is in January, February, June, August, September, October, and December. Analyzing the reasons, this may be due to the relatively easy work tasks for tourists during this time period, while students are in the holiday phase.

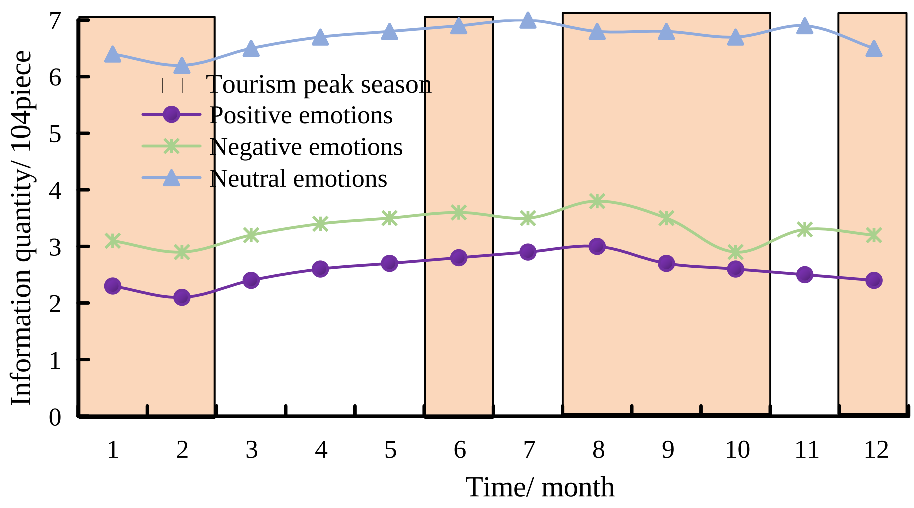


Figure 6: Statistical situation of research data in different research areas

Fig. 7 shows the statistical data of male and female tourists in different months. The number of information pieces for male tourists is significantly less than that for female tourists, accounting for only about 1/4. This is partly due to the fact that female tourists enjoy recording their travel experiences and experiences, and there is also an objective phenomenon of a higher number of female tourists.

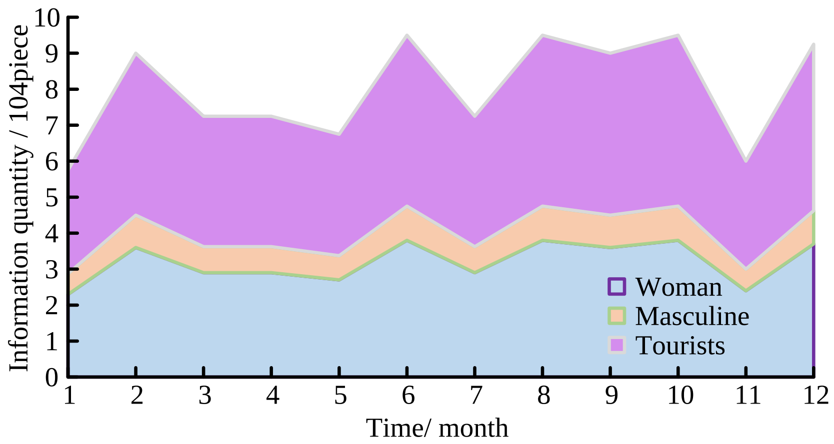


Figure 7: Statistics on male and female tourists in different months

This study analyzes the performance of TPRA, comparing algorithms such as Probabilistic Matrix Factorization with User and Item relationships (PMFUI) algorithm, SoRec method, and Probabilistic Matrix Factorization (PMF). Simultaneously, four benchmark testing functions are utilized to accurately evaluate the algorithm performance. Among them, the f1 function is used to evaluate the performance of the algorithm in recommending paths based on user preferences. User preferences for paths are mainly influenced by the number and type of attractions in the path. The f2 function simulates the user's preferences for time and distance in the travel process, with the objective of evaluating the performance of the algorithm when considering time and distance factors. The f3 function simulates the user's preference for the correlation between various attractions in the path, used to evaluate the performance of the algorithm when considering the correlation between attractions. The f4 function simulates the comprehensive evaluation of the overall quality of the path by users during the tourism process. Figures 8 (a) - 8 (d) show the performance of different TPRAs under different benchmark test functions and runs, respectively. Overall, for the four benchmark functions, PSO-AC has a longer runtime under the same number of runs. Under the four benchmark test functions f1, f2, f3, and f4, when the maximum runs are reached, the running time is 36.58 s, 62.96 s, 90.59 s, and 64.26 s, respectively. Taking Fig. 8 (a) as an example, when the run reaches 30, the running times of PMF, SoRec, PMFUI, and PSO-AC are 15.26 s, 18.19 s, 23.49 s, and 36.27 s, respectively. Compared to the other three types of TPRA, the increase in operating time of PSO-AC is 114.26%, 101.32%, and 58.29%, respectively. Therefore, under the same number of runs, the PSO-AC algorithm has a longer running time, but the difference in running time is significant.

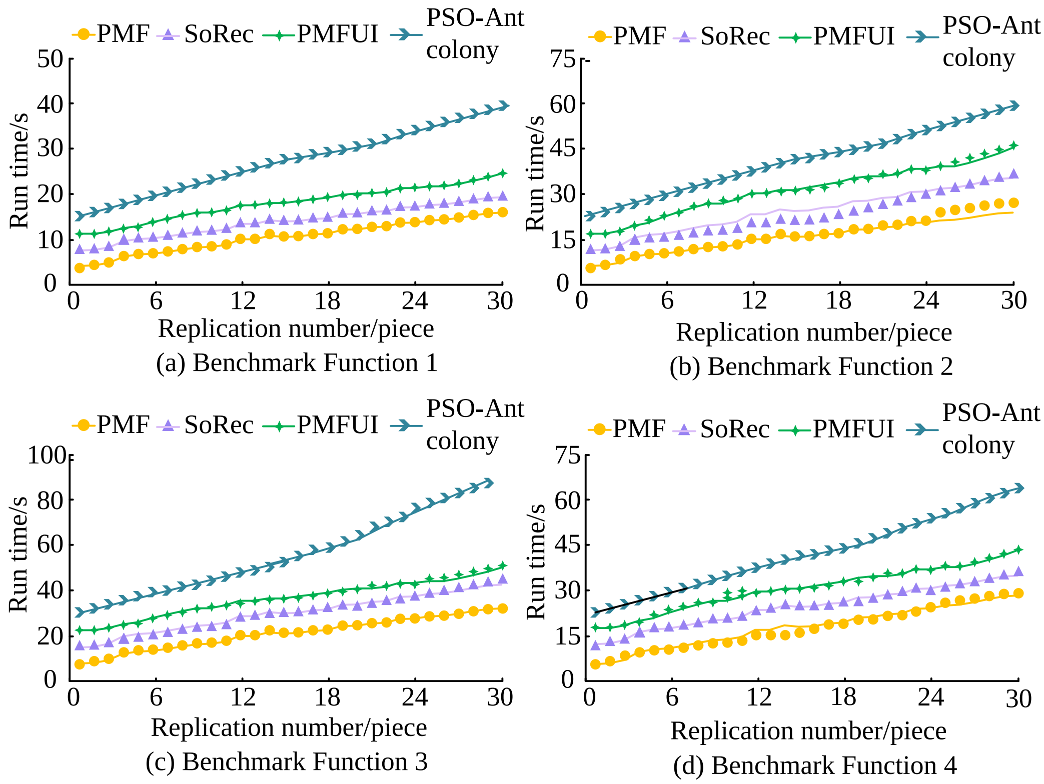


Figure 8: Performance of different TPRAs under different benchmark testing functions and running times

Figures 9 (a) - 9 (d) show the accuracy and error values of four different TPRAs under different benchmark test functions. Overall, the error of TPRA decreases continuously with the increase of data size, and tends to stabilize when the data size reaches around 5, while the accuracy shows the opposite trend of change. In Fig. 4 (a), the convergence errors of PPMF, SoRec, PMFUI, and PSO-AC are 0.410, 0.309, 0.256, and 0.145, respectively, with accuracy of 91.26%, 93.56%, 95.21%, and 98.26%, respectively. In Fig. 4 (b), the convergence errors of PMF, SoRec, PMFUI, and PSO-AC are 0.512, 0.465, 0.326, and 0.198, with accuracy of 87.25%, 89.65%, 92.26%, and 93.58%. Therefore, the accuracy of PMF, SoRec, PMFUI, and PSO-AC is higher than that of other optimization algorithms, while the error is lower.

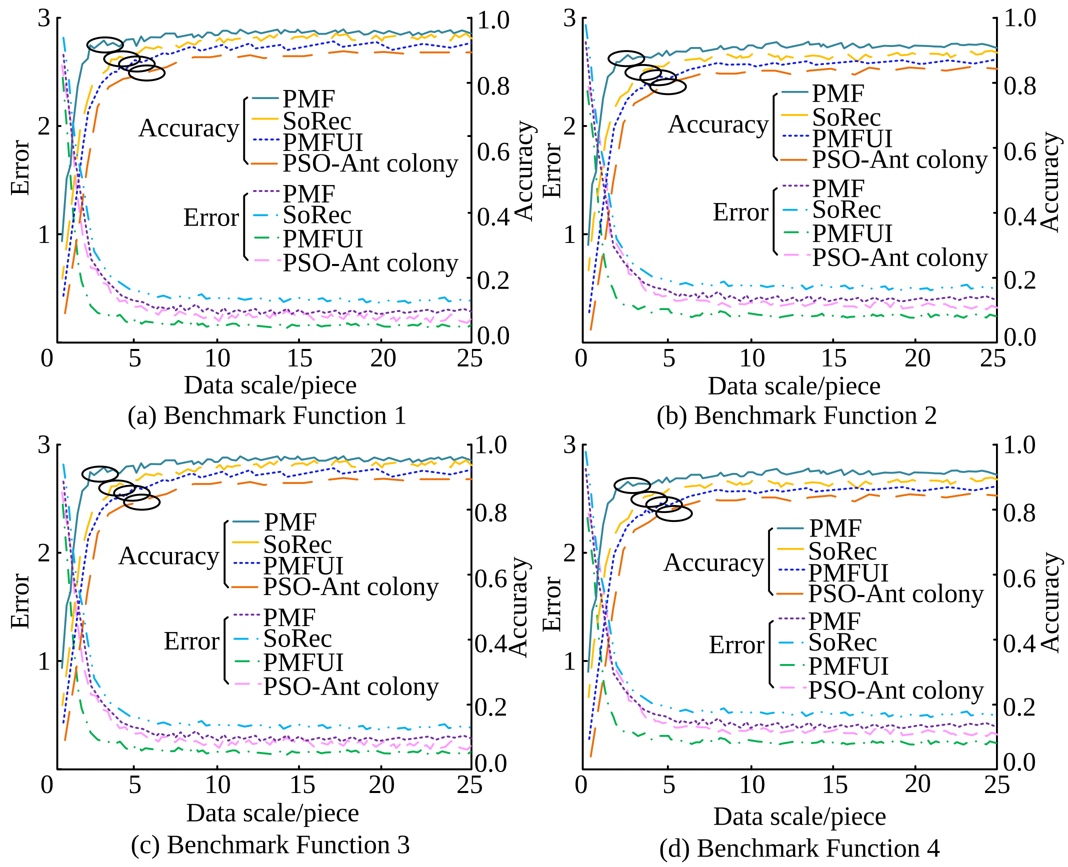
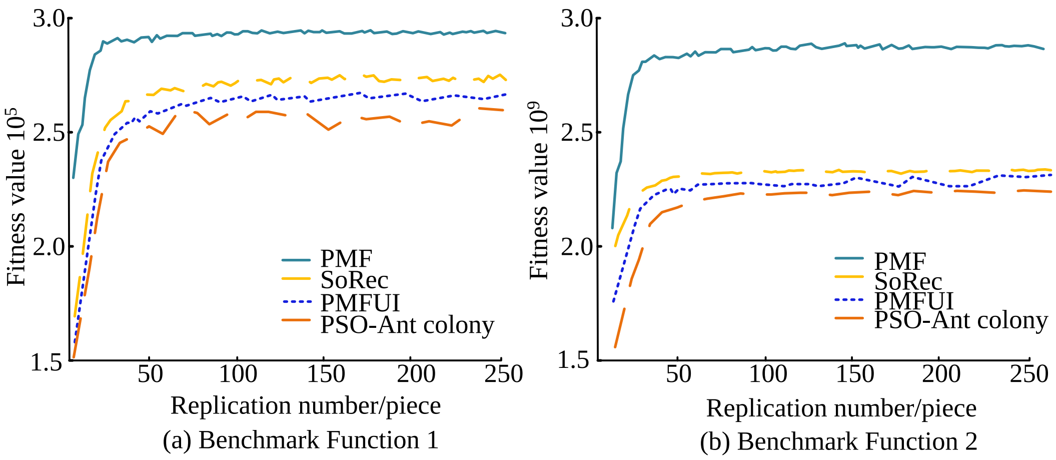


Figure 9: Accuracy and error values of four different TPRAs under different benchmark test functions

Figures 10 (a) - 10 (d) show the fitness values of four different TPRAs under different benchmark test functions. Compared to other algorithms, PSO-AC has better optimal fitness and average fitness. This is because the computational complexity of AC does not overly value the dimensions of optimization problems, and it has high search accuracy and strong generalization ability. The introduction of Max/Min-M and EAMM in AC can effectively solve the convergence speed, making it more conducive to obtaining the global optimal solution. PSO-AC is stable and efficient, and the average fitness of the two different TPRAs, SoRec and PMFUI, is better than PMF.



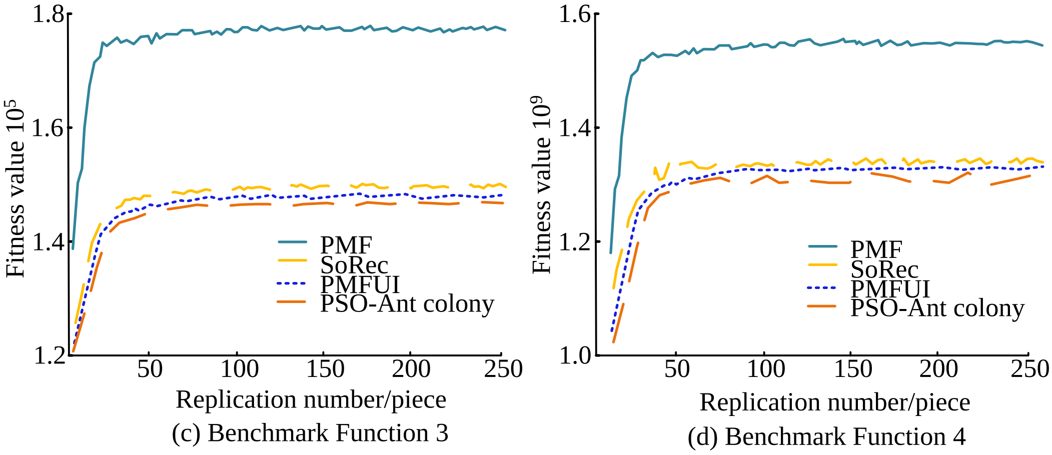


Figure 10: Fitness values of four different TPRAs under different benchmark test functions

3.2 Optimization effect of PSO-AC

To analyze the optimization effect of PSO-AC, this study explores the fitness value and recommendation error. Figures 11 (a) and 11 (b) show the relationship between fitness values, running time, and recommendation error at different iterations. PSO-AC reaches its optimal value after approximately 600 iterations, with a running time of 7.452 seconds and an optimal fitness value of 0.475. The iteration error of other TPRAs is higher, and the running time is also around 7.5 seconds. For recommendation accuracy, PSO-AC has lower recommendation errors under different running times, with TPR error values ranging from 0.005 to 0.089, while PMF has the lowest recommendation accuracy, with TPR error from 0.154 to 0.268. The recommendation accuracy of SocRec and PMFUI TPRA is in the middle, with values from 0.098 to 0.189 and 0.078 to 0.105.

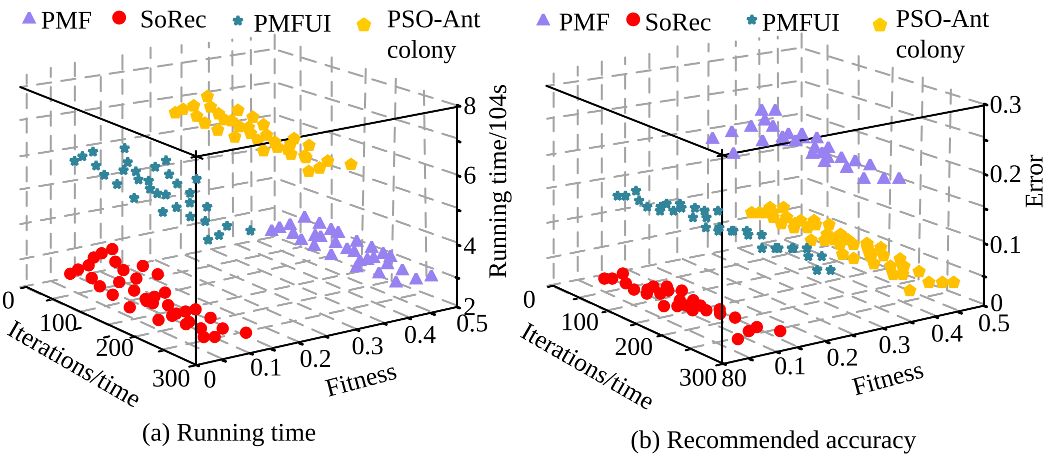
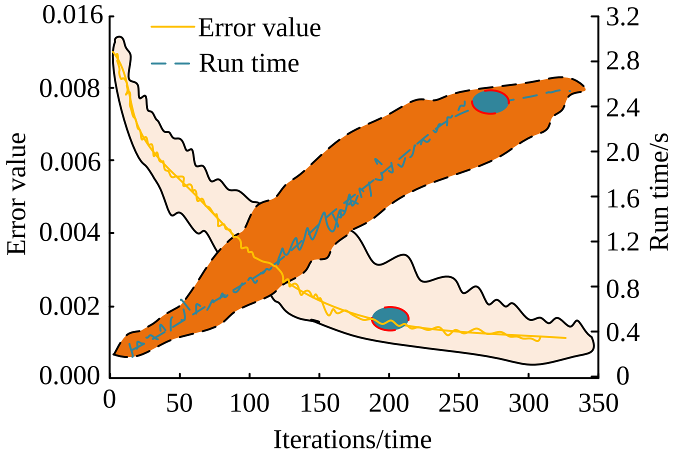
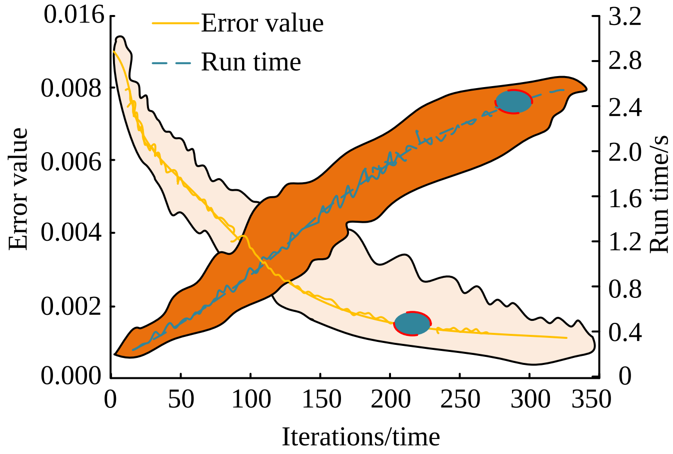


Figure 11: The relationship between fitness value, running time, and recommendation error under different iteration times

To further validate the effectiveness of the improved AC method, this paper compares it with the pre-improved TPRA. Figures 12 (a) and 12 (b) show the convergence values and running times. In Fig. 12 (a), the error values and iteration times of the two algorithms are inversely related, and the final convergence times are about 200 and 210. The stable error value of PSO-AC is 0.014, which is 28.36% higher than that of a single algorithm. In Fig. 12 (b), the stable values are 2.9s and 2.8 s, indicating that the iterative process of PSO-AC is still within a reasonable numerical range and there is not much difference in values compared to the improved TPRA. Therefore, PSO-AC is excellent in optimizing time and error values.



(a) Improve JAYA



(b) Adaptive JAVA

Figure 12: Convergence values and runtime of two algorithms

Fig. 13 shows the confusion matrix results of PSO-AC before and after improvement on the training set. 1-5 refers to five different themes. Fig. 13 (a) shows that the TPR accuracy for themes 1 and 5 is 71.25% and 72.26%, respectively, while the TPR accuracy for the other three themes is around 70%. The confusion matrix results in Fig. 13 (b) show that the TPR accuracy for themes 1 and 5 is 81.25% and 84.26%, respectively, and the TPR accuracy for the other three themes is also above 75%. The improved TPR accuracy of PSO-AC is significantly better, with an improvement rate of about 15%. Common types of TPR errors include recommended travel paths that do not match the user's actual interest topics, and recommended travel paths that fail to consider the user's personalized needs. The reason for the former is that the algorithm did not correctly recognize the user's preferences, or the user's interest topics were misunderstood during the recommendation process. The latter may be due to a lack of sufficient understanding of user preferences, or errors in the data collection and analysis process. The improved PSO-AC algorithm has significant advantages in both aspects.

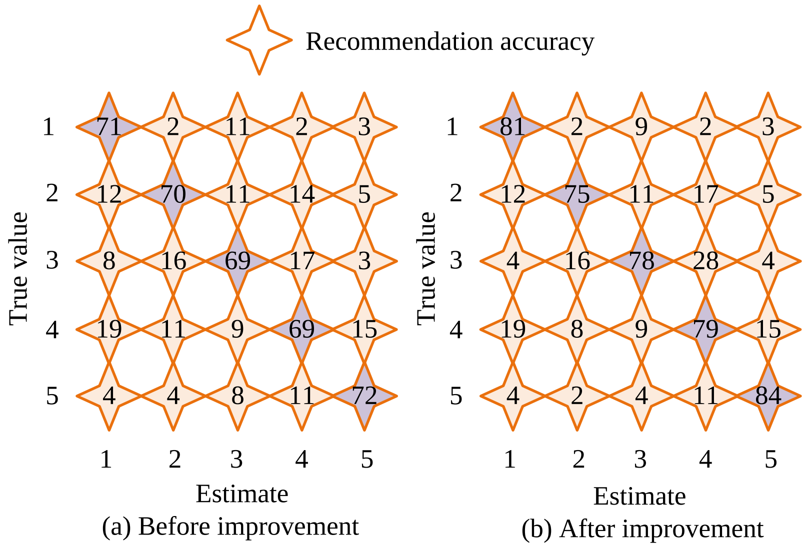


Figure 13: The confusion matrix results of PSO-AC improved TPRA on the training set

Fig. 14 shows the confusion matrix results of PSO-AC on the training and testing sets. The PSO-AC in the training set showed that the TPR accuracy for themes 1 and 5 was 81.25% and 84.26%, respectively, and the TPR accuracy for the other three themes also reached over 75%. This may be due to the fact that these two themes are favored by tourists and there is a large amount of textual information for evaluation. In the test set, the TPR of PSO-AC in themes 1 and 5 were 85.26% and 87.25%, respectively, and the TPR accuracy of the other three themes reached over 78%. Therefore, the accuracy of the proposed TPRA is relatively high, and the accuracy of the TPRA improved by the AC improvement scheme has been improved. This recommendation method has certain feasibility in TPR.

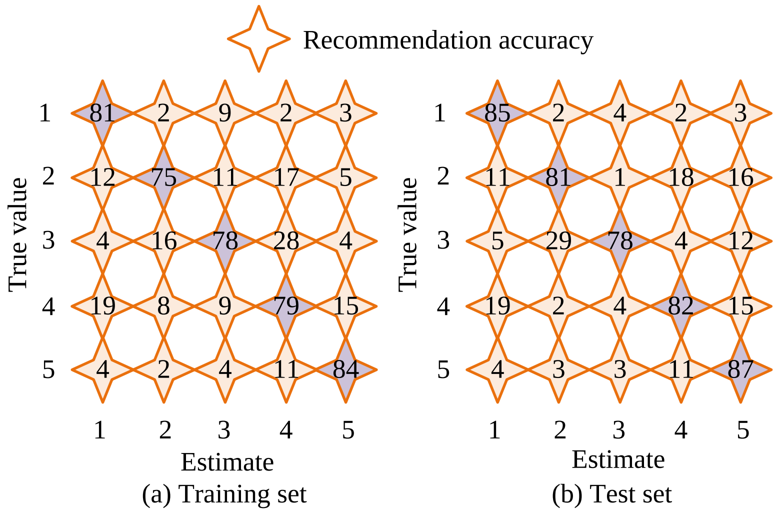


Figure 14: Confusion matrix results of PSO-AC TPRA on test and training sets

To further evaluate the performance of the proposed algorithm, accuracy, recall, and F1 value are used as performance evaluation indicators, and compared with the improved TPRA. The performance evaluation indicators of each algorithm under different iteration times are shown in Fig. 15. As shown in Fig. 15 (a), the highest accuracy of the pre improved TPRA is only 67.91%, while the accuracy of the improved TPRA is as high as 83.64%. As shown in Fig. 15 (b), the improved TPRA has a recall rate of 77.91%, which is 33.01% higher than before the improvement. From Fig. 15 (c), the F1 value of the improved TPRA is as high as 83.41%, which is significantly higher than the original recommendation algorithm. The improved PSO-AC TPRA has significant performance advantages.

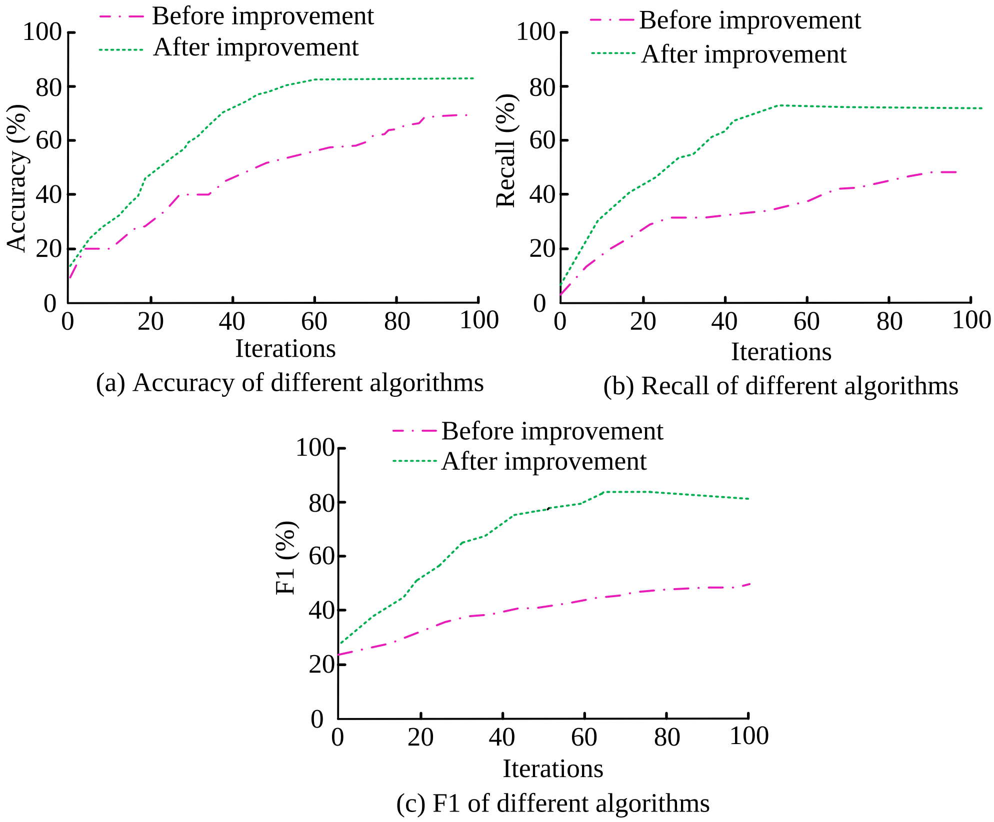


Figure 15: Performance evaluation indicators of various algorithms under different iteration times

4 Discussion

In terms of runtime, the PSO-AC TPRA reaches its optimal value after approximately 600 iterations, with a runtime of 7.452 seconds. The recommendation time of the personalized travel route intelligent recommendation technology designed by Luo et al. [12] with improved discrete PSO is within 0.8 seconds, slightly higher than the method proposed in the study. The reason is that the PSO-AC algorithm has a lower algorithm complexity and uses an efficient heuristic search strategy, which can find the optimal solution in a shorter time. In terms of recommendation accuracy, the proposed PSO-AC TPRA has lower recommendation errors under different running times. The range of error values for TPR is 0.005-0.089, while the PMF TPRA has the lowest recommendation accuracy, with the range of error values for TPRs being 0.154-0.268. The recommendation accuracy of the SocRec and PMFUI is in the middle, with values ranging from 0.098 to 0.189 and 0.078 to 0.105, respectively. The recommendation error rate of the optimal PSO and optimal ant recommendation system proposed by Qamar et al. [11] is less than 5%. Compared with the proposed PSO-AC, its recommendation error rate is at a disadvantage. The reason is that the PSO-AC algorithm has strong global search ability and strong adaptability, which reduces the recommendation error rate. In the confusion matrix results of five different themes, the accuracy of TPRs for themes 1 and 5 is 81.25% and 84.26%, respectively, and the accuracy of TPRs for the other three themes is also above 75%. The accuracy of the probability category recommendation algorithm proposed by Javed Awan et al. [9] is only 71.62%. Researchers such as Beed R [10] proposed the artificial bee colony optimization algorithm and PSO algorithm to obtain a maximum TPR accuracy of 81.51%. Compared with the maximum recommendation accuracy of the PSO-AC, it has a higher recommendation accuracy, which is 12.64% and 2.75% lower, respectively. The reason is that PSO algorithm excels in global search, while AC algorithm performs well in local search and information transmission. The two complement each other, allowing PSO-AC algorithm to achieve higher accuracy in TPR. In addition to its applicability to tourism itinerary planning, the PSO-AC can be extended to other fields, such as logistics distribution and transportation planning. This enables the provision of efficient path planning and optimization solutions for related fields.

5 Conclusion

In response to the shortcomings of existing personalized recommendation algorithms in user preference information data mining on social media platforms, as well as the problems in personalized recommendation scheme analysis, this study designed a TPRA that combined PSO and AC. Meanwhile, it optimized AC using Max/Min-M and EAMM. The error of TPRA decreased continuously with the increase of data size, and tended to stabilize when the data size reached about 5, while the accuracy showed the opposite trend of change. For benchmark function f1, the convergence errors of PPMF, SoRec, PMFUI, and PSO-AC were 0.410, 0.309, 0.256, and 0.145, respectively, and the accuracy was 91.26%, 93.56%, 95.21%, and 98.26%, respectively. The optimal fitness and average fitness of PSO-AC were better. PSO-AC reached its optimal value after approximately 600 iterations, with a running time of 7.452 seconds and an optimal fitness value of 0.475. The iteration error of other TPRA was higher, and the running time was also around 7.5 seconds. The stable error value of TPRA improved by AC was 0.014, which is 28.36% higher than the original algorithm. The stable values of the TPRA algorithms before and after AC improvement were 2.9s and 2.8s, respectively. The confusion matrix results of TPRA before and after PSO-AC improvement showed that the TPR accuracy of topics 1 and 5 was the highest, but the improved values were higher, with an improvement rate of 15%, and the TPR accuracy of the other three topics was around 70%. This TPRA not only effectively shortens the recommendation time for the best route, but also enables personalized and accurate recommendation of recommended routes, which is superior to other TPRAs currently available. However, there are still certain limitations. The data collected in the experiment is only a portion of the fixed area of the user's travel trajectory, and the sparsity of the data itself will affect the accuracy of the recommendation, which may reduce the accuracy of the results. In future research, it is necessary to further optimize the PSO-AC tourism recommendation algorithm by introducing more user features and preference information to improve recommendation accuracy. In the case of larger datasets, it is necessary to consider the time and spatial complexity of the algorithm. By optimizing the internal data structure and computational logic of the algorithm, as well as utilizing parallel computing to improve the efficiency of the algorithm, it is possible to enhance the overall performance of the algorithm. Considering the differences in different geographical regions, it is necessary to consider using region specific parameter settings or model adjustments. In addition, it is necessary to apply the PSO-AC algorithm to practical applications. It is recommended to recruit some actual users, including travel enthusiasts and ordinary travelers, and provide them with PSO-AC recommended travel paths, collect their user experience and satisfaction feedback, and verify the practicality of the algorithm.

References

[1] X. Zhao, X. Yan, A. Yu, and P. Van Hentenryck, "Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models," Travel Behaviour and Society, vol. 20, pp. 22-35, 2020. https://doi.org/10.1016/j.tbs.2020.02.003

[2] H. Jain, A. Khunteta, and S. Srivastava, "Churn prediction in telecommunication using logistic regression and logit boost," Procedia Computer Science, vol. 167, pp. 101-112, 2020. https://doi.org/10.1016/j.procs.2020.03.187

[3] S. Nusinovici, Y. C. Tham, M. Y. C. Yan, D. S. W. Ting, J. Li, C. Sabanayagam, and C. Y. Cheng, "Logistic regression was as good as machine learning for predicting major chronic diseases," Journal of Clinical Epidemiology, vol. 122, no. 7, pp. 56-69, 2020. https://doi.org/10.1016/j.jclinepi.2020.03.002

[4] P. Preethi and H. R. Mamatha, "Region-based convolutional neural network for segmenting text in epigraphical images," Artificial Intelligence and Applications, vol. 1, no. 2, pp. 119-127, 2023. https://doi.org/10.47852/bonviewAIA2202293

[5] K. Bhosle and V. Musande, "Evaluation of deep learning CNN model for recognition of devanagari digit," Artificial Intelligence and Applications, vol. 1, no. 2, pp. 114-118, 2023. https://doi.org/10.47852/bonviewAIA3202441

[6] A. Abba Haruna, L. J. Muhammad, and M. Abubakar, "Novel thermal-aware green scheduling in grid environment," Artificial Intelligence and Applications, vol. 1, no. 4, pp. 244-251, 2022. https://doi.org/10.47852/bonviewAIA2202332

[7] A. Jamal, M. Zahid, M. Tauhidur Rahman, H. M. Al-Ahmadi, M. Almoshaogeh, D. Farooq, and M. Ahmad, "Injury severity prediction of traffic crashes with ensemble machine learning techniques: A comparative study," International Journal of Injury Control and Safety Promotion, vol. 28, no. 4, pp. 408-427, 2021. https://doi.org/10.1080/17457300.2021.1928233

[8] Y. Huang, W. Li, F. Macheret, R. A. Gabriel, and L. Ohno-Machado, "A tutorial on calibration measurements and calibration models for clinical prediction models," Journal of the American Medical Informatics Association, vol. 27, no. 4, pp. 621-633, 2020. https://doi.org/10.1093/jamia/ocz228

[9] M. Javed Awan, M. S. Mohd Rahim, H. Nobanee, A. Munawar, A. Yasin, and A. M. Zain, "Social media and stock market prediction: a big data approach," Computers, Materials & Continua, vol. 67, no. 2, pp. 2569-2583, 2021. https://doi.org/10.32604/CMC.2021.014253

[10] R. Beed, A. Roy, S. Sarkar, and D. Bhattacharya, "A

hybrid multi-objective tour route optimization algorithm based on particle swarm optimization and artificial bee colony optimization," Computational Intelligence, vol. 36, no. 3, pp. 884-909, 2020. https://doi.org/10.1111/coin.12276

[11] M. S. Qamar, S. Tu, F. Ali, A. Armghan, M. F. Munir, F. Alenezi, and N. Alnaim, "Improvement of traveling salesman problem solution using hybrid algorithm based on best-worst ant system and particle swarm optimization," Applied Sciences, vol. 11, no. 11, pp. 4780-4795, 2021. https://doi.org/10.3390/app11114780

[12] J. Luo and X. Duan, "Intelligent recommendation of personalised tourist routes based on improved discrete particle swarm," International Journal of Computational Science and Engineering, vol. 25, no. 6, pp. 598-606, 2022. https://doi.org/10.1504/IJCSE.2022.127189

[13] S. González-Bailón and Y. Lelkes, "Do social media undermine social cohesion? A critical review," Social Issues and Policy Review, vol. 17, no. 1, pp. 155-180, 2023. https://doi.org/10.1111/sipr.12091

[14] H. Zhang, F. Liu, L. Ma, and Z. Zhang, "A hybrid heuristic based on a particle swarm algorithm to solve the capacitated location-routing problem with fuzzy demands," IEEE Access, vol. 8, pp. 153671-153691, 2020. https://doi.org/10.1109/ACCESS.2020.3018490

[15] X. Zhou, H. Ma, J. Gu, H. Chen, and W. Deng, "Parameter adaptation-based ant colony optimization with dynamic hybrid mechanism," Engineering Applications of Artificial Intelligence, vol. 114, no. 9, pp. 105139-105164, 2022. https://doi.org/10.1016/j.engappai.2022.105139

[16] M. S. Kumar and J. Prabhu, "A hybrid model collaborative movie recommendation system using K-means clustering with ant colony optimisation," Int J Int Te Sec Trans, vol. 10, no. 3, pp. 337-354, 2020. https://doi.org/10.1504/IJITST.2020.107079

[17] P. Stodola, K. Michenka, J. Nohel, and M. Rybanský, "Hybrid algorithm based on ant colony optimization and simulated annealing applied to the dynamic traveling salesman problem," Entropy, vol. 22, no. 8, pp. 884-897, 2020. https://doi.org/10.3390/E22080884

[18] M. Hamim, I. E. Moudden, M. D. Pant, H. Moutachaouik, and M. Hain, "A hybrid gene selection strategy based on fisher and ant colony optimization algorithm for breast cancer classification," International Journal of Online & Biomedical Engineering, vol. 2021, no. 2, pp. 148-163, 2021. https://doi.org/10.3991/IJOE.V17I02.19889

[19] M. Abdolhosseinzadeh and M. M. Alipour, "Design of experiment for tuning parameters of an ant colony optimization method for the constrained shortest Hamiltonian path problem in the grid networks," Numerical Algebra, Control and Optimization, vol. 11, no. 2, pp. 321-332, 2021. https://doi.org/10.3934/naco.2020028.

[20] R. Syah, M. K. M. Nasution, E. B. Nababan, and S. Efendi, "Sensitivity of shortest distance search in the ant colony algorithm with varying normalized distance formulas," TELKOMNIKA, vol. 19, no. 4, pp. 1251-1259, 2021. https://doi.org/10.12928/TELKOMNIKA.v19i4.18872

[21] J. Luo and X. Duan, "Intelligent recommendation of personalised tourist routes based on improved discrete particle swarm," International Journal of Computational Science and Engineering, vol. 25, no. 6, pp. 598-606, 2022. https://doi.org/10.1504/IJCSE.2022.127189

[22] P. S. Priya and N. N. M. Rao, "Potential of swarm intelligence-based tour and travel recommendations systems," International Journal of Intelligence Systems Applications in Engineering, vol. 11, no. 10, pp. 763-772, 2023. https://doi.org/10.1504/IJCSE.16.08.2023

[23] X. Liu, D. Zhang, T. Zhang, J. Zhang, and J. Wang, "A new path plan method based on hybrid algorithm of reinforcement learning and particle swarm optimization," Engineering Computations, vol. 39, no. 3, pp. 993-1019, 2022. https://doi.org/10.1108/EC-09-2020-0500

[24] Y. Li, "Digital tourism recommendation and route planning model design based on ripplenet and improved GA," Informatica, vol. 48, no. 3, pp. 133-152, 2024. https://doi.org/10.31449/inf.v48i10.5790

[25] R. A. Pop, Z. Săplăcan, D. C. Dabija, and M. A. Alt, "The impact of social media influencers on travel decisions: The role of trust in consumer decision journey," Current Issues in Tourism, vol. 25, no. 5, pp. 823-843, 2022. https://doi.org/10.1080/13683500.2021.1895729

[26] X. Liu, D. Zhang, T. Zhang, J. Zhang, and J. Wang, "A new path plan method based on hybrid algorithm of reinforcement learning and particle swarm optimization," Engineering Computations, vol. 39, no. 3, pp. 993-1019, 2022. https://doi.org/10.1108/EC-09-2020-0500

[27] N. Muñoz-Izquierdo, E. K. Laitinen, M. D. M. Camacho-Miñano, and D. Pascual-Ezama, "Does audit report information improve financial distress prediction over Altman's traditional Z-Score model?" Journal of International Financial Management & Accounting, vol. 31, no. 1, pp. 65-97, 2020. https://doi.org/10.1111/jifm.12110

[28] H. K. Andi, "An accurate bitcoin price prediction using logistic regression with LSTM machine learning model," Journal of Soft Computing Paradigm, vol. 3, no. 3, pp. 205-217, 2021. https://doi.org/10.36548/jscp.2021.3.006

[29] J. Sun, H. Fujita, Y. Zheng and W. Ai, "Multi-class financial distress prediction based on support vector

machines integrated with the decomposition and fusion methods," Information Sciences, vol. 559, no. 6, pp. 153-170, 2021. https://doi.org/10.1016/j.ins.2021.01.059

[30] R. Khera, J. Haimovich, N. C. Hurley, R. McNamara, J. A. Spertus, N. Desai, and H. M. Krumholz, "Use of machine learning models to predict death after acute myocardial infarction," JAMA Cardiology, vol. 6, no. 6, pp. 633-641, 2021. https://doi.org/10.1001/jamacardio.2021.0122