Journal of Nonlinear and Convex Analysis
Volume 26, Number 6, 2025, 1615–1628

Yokohama Publishers
ISSN 1880-5221 ONLINE JOURNAL

RESEARCH ON SUPPORT VECTOR MACHINE-BASED URBAN RECREATIONAL PUBLIC TRANSPORTATION LINE DISCRIMINATION

XIAOXI JI AND ANMIN HUANG*

Abstract. The objective of urban recreational transportation networks in supporting urban recreational spatial distribution, layout of urban leisure facilities, urban planning, and urban recreation service level is significant. In addition to being a typical leisure city, Xiamen is a significant tourism city in China. This article is composed of information gathered from multiple sources, including "public transportation IC card data", "GPS arrival data of public transportation", and "static route station information of public transportation" in Xiamen City. Here, we build a support vector machine model and evaluate KNN and DTs approaches using ticket price, first bus time, maximum departure frequency, final bus time, lowest departure frequency, and the number of stations in the buffer zone as eigenvalues. The findings revealed that the SVM model beat the other two approaches with 96.1% accuracy, 96.4% AUC, 87.5% KNN model accuracy, 92.1% AUC, and 75.7% DTs model accuracy. The classification method of this study can quickly and effectively identify recreational routes in cities, helping urban managers optimize the recreational public transportation network, exploring urban recreational people's time and space travel rules, and providing empirical basis for urban recreational spatial layout and functional zoning.

1. Introduction

An increasingly significant aim of urban development is the development of urban comprehensive public transport, particularly against the backdrop of green travel and low-carbon travel. Urban public transportation systems have broadened their service aims to include not just city dwellers but also urban recreational visitors, including international tourists, as a result of the advent of mass tourism. Consequently, there is a growing need for study into the best ways to construct public transportation networks in cities that are accessible to both locals and tourists, conducive to leisure activities, make efficient use of available resources, and promote intelligent city planning.

The advent of cloud computing and the proliferation of the Internet have brought big data to the forefront, and the field is seeing rapid growth in both research

2020 *Mathematics Subject Classification.* 00A06, 65Z05.

Key words and phrases. Multi-source data, urban recreation, public transport lines, support vector machine, Xiamen.

This work was supported in part by the National Social Science Fund of China under Grant 17BGJ043, in part by the Huaqiao University Postgraduate Research Innovation Ability Cultivation Program under Grant 17011121003,Tertiary Education Scientific research project of Guangzhou Municipal Education Bureau 202235343, and "Basic and Applied Basic Research Topic for 2025 (Young Doctoral 'Setting Sail' Project) Number:SL2024A04J01462".

*Corresponding author.

and practical applications. The methodologies used for collecting and analyzing transportation data provide the basis of transportation network research. Urban planning, transportation planning, public services, urban tourism, and other related domains might benefit from the analytical resources made possible by mining this data set for insights into urban transportation developments [13].

Location-based services (LBS) such as mobile communication (GSM), global positioning system (GPS), social network (SNS), and wireless broadband hotspot (Wi-Fi) currently make up the majority of multi-source data in multiple industries. Technique of data records and location service provide records of bus smart-card, flight, credit card, Weibo and mobile phone, etc. [8] data collection of bus IC card is simultaneously completed when passengers getting on and off bus swipe the cards. Due to the convenience of transaction and the accuracy and reliability of recorded data, the method has gradually become a widely used means to analyze the intensity of urban activities, spatial and temporal distribution characteristics and the travel trajectory of public transport passenger flow. Current research includes the impact of urban road planning on people's travel, urban population travel characteristics like occupation and residence [14], urban public transport passenger travel characteristics [7], and the impact of public transport on Urban Recreational Spatial Pattern. [21] However, few studies have examined the distribution of urban recreation space, recreational amenities, and urban planning, an essential matter. Urban tourism and host-guest sharing are quickly transforming the urban population structure. Thus, the urban leisure population expanded from local recreation to comprehensive local recreation and international visitors. Real-time and accurate recreational distribution data could inspire urban planners to design the distribution of urban recreational space in the future, as well as validate prior planning. Traditional research typically fails urban planners and researchers owing to low data volume and inability to update in real time due to lack of data gathering and statistical methods. Therefore, with the development of big data mining technology, using the activity track of a large number of urban leisure people from multiple sources, it has gradually become a hot spot of urban tourism research in the future. Hence, employing the activity track of multi-source and significant urban leisure people could become the future hotspot of urban tourism research as big data mining technology developments. J.B. (2021) propose a methodology to identify tourist routes that integrate the most important Points Of Interest in a region taking up as criteria profile characteristics in common between the sites evaluated using clustering techniques. [5] Manal Addoun (2023) using semi-structured interviews and questionnaires for the design of these routes to analyze the cultural tourist routes. [1]

With the development of machine learning theory and the development of technology, numerous classification algorithms have emerged, such as logistic regression (LR), decision tree (DT), Bayes statistics (BS), K-Nearest Neighbors (KNN), neural network (NN), and support vector machine SVM. [24] In particular, support vector machines (SVMs) have attracted a lot of interest and study from researchers because of the problems they effectively resolve, including under-learning, overfitting, and the curse of dimensionality. SVMs are powerful classification algorithms that have solid mathematical and statistical theoretical underpinnings in previous studies. The method is one of the most effective in machine learning and pattern recognition, having been used extensively and developed in areas including face recognition, image processing, and information retrieval. [6] In the current research, support vector machine is mainly used in the prediction of tourism demand. Due to the characteristics of short time, complex influencing factors and difficulty in obtaining tourism demand prediction statistical data. The experiments conducted by Yin Ying et al (2004) showed that the support vector machine theory and model have great application potential in tourism demand forecasting. [22] In view of the characteristics of tourism factors that are difficult to determine and small sample size in tourism forecasting, the comparative research results of Li Zhilong et al (2010) show that the SVR model is better than the BPNN model. [11] Yu et al showed that: using LS-SVM method for medium-term prediction of tourist flow in scenic areas, the predicted performance indexes are significantly better than Back Propagation Neural Network (BP), Autoregressive Integrated Moving Average Model (X-12-ARIMA), Empirical Mode Decomposition (EMD) and LS-SVM combined prediction method. [23] This paper presents the establishment of a support vector machine (SVM) based classification algorithm model. The model is trained using eigenvalue extraction, solves the SVM model, and identifies recreational public transport routes based on data of bus stops, 397 lines, and a continuous month-long resident travel trajectory within a 1000-meter buffer zone surrounding 40 recreational sites in Xiamen. Ultimately, 159 Xiamen leisure paths are found. The identification technique in this study is transferable to other domestic and international cities. More communities could benefit from this study by using it as a guide when they develop and optimize their recreational transportation routes.

2. Indicators and methods

SVM, KNN, and DT are widely popular in many applications, and this study chooses SVM, KNN, and DT for comparative research [2, 9, 10, 12, 16, 17, 19]. Existing research utilizes network analysis tools to model and compare the travel time of public transportation within the existing Melbourne public transportation network. These studies introduce an innovative approach where researchers can obtain rich information about public transportation networks, including real-time travel times, passenger demand patterns, and information on tourist destination stations. This rich dataset provides a solid foundation for extracting meaningful feature values, helping to establish classification models more accurately and conducting more scientific analysis of the results.In this study, the use of big data analysis tools represents a significant advancement in the field of public transportation line classification analysis. It can provide a more comprehensive understanding of the different attributes of transportation systems and valuable insights for traffic planners and policy makers to optimize and improve existing public transportation networks.

The research methodology employed in this study is capable of swiftly and efficiently recognizing recreational routes within urban areas, thereby assisting urban administrators in refining the recreational public transportation system, offering a solid empirical foundation for the spatial arrangement and functional zoning of urban recreational areas.

2.1. **Indicator Description.** The term "ticket price" refers to the "transaction amount" field in the passenger card data. In this case, the transaction amount of the buses is 0.8 yuan, 1.6 yuan, and 5 yuan; the first two represent regular bus routes, while the last means using the tourist special line. The expense of the ticket could be utilized as a distinctive value of recreational routes by marking the training set and training model, as it clearly distinguishes normal buses from tourist special lines. The number of bus lines is denoted by the line number, which is utilized to differentiate between lines. In the line set database, a line has different number of stations, which together constitute a bus line. Concurrently, N lines will pass through a given station; hence, it is important to assign responsibility for passenger flow based on the relationship between line number and station name.

One of the information gathered from the data set of passenger swipe cards is the initial bus time. This number represents, in seconds, the first passenger to board an identified bus route throughout the day. The data set of passengers swiping cards contains information such as the last bus time. As a number of seconds, it shows the time that the very last person got on a certain line of buses that day. The minimum departure frequency is one of the information collected from the static route information collection, and it relates to the time interval between the departures of each bus on the bus route, establishing the departure frequency of this route. A feature that has been chosen for inclusion in this article is the minimum departure interval duration.

The time interval of each bus departure in the bus route, which represents the departure frequency of this route, is one of the types of information derived from the static route information dataset of the highest departure frequency. In this article, the maximum departure interval time is one of the aspects that has been chosen with consideration. There are three different kinds of recreational destinations inside the 1000 meter buffer zone: stations, commercial malls, and scenic sites. The number of bus stops within 1000 meters of the geographical position coordinate linear radius, and the number of stations in the buffer zone of the line where the station is situated, the more evident the recreation feature of the line.

2.2. **SVM Classification Method.** The SVM algorithm (Support Vector Machine), a binary classification technique based on contemporary statistical theory, differs from classic statistical approaches. The measurement of probabilities and the law of big numbers are essentially insignificant. In order to solve nonlinear, high-dimensional, and small sample problems, it offers several special benefits along with excellent accuracy. The article features a line with a total sample size of 397, which, in terms of line attributes, corresponds to a small sample and high latitude. As a result, the chosen SVM support vector machine model has an excellent application. The two kinds of samples in Figure 2 are represented by the solid and hollow points, respectively, and by the classed hyperplane H between them. The samples H1 and H2 are those that are closest to the classification surface and parallel to the classification plane. The classification interval is the distance between them, while H1 and H2 stand for general routes and recreation routes, respectively. [3, 18, 20]

The sample set $(x_i, y_i), i = 1, 2, \ldots, n,\ x \in R^d, y \in \{+1, -1\}$ that is linearly separable areas category label $(x_i, y_i), i = 1, 2, \ldots, n,\ x \in R^d, y \in \{+1, -1\}$. The

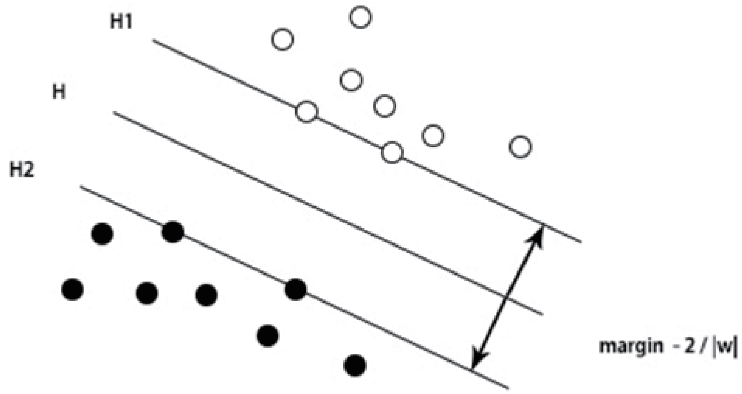

FIGURE 1. Linear separable support vector machine

general form of linear discriminant function in -dimensional space is symbolic function. The general form of linear discriminant function in -dimensional space is $g(x) = w \cdot x + bg(x) = w \cdot x + b$ (w is y of Hilbert and b is weight offset vector in space. The classification interval in the sample set is equal $2/\|w\|$ to $2/\|w\|$, so maximizing the interval is equivalent to minimizing $\|w\|\,\|w\|^2\,\|w\|$ or $\|w\|^2$. To require classification to correctly classify all samples is to require it to satisfy the objective function:

$$L = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \alpha_i \left\{ y_i \left(w x_i + b \right) - 1 \right\} \tag{2.1}$$

ai is the constraint condition of Lagrange operator. The maximum and minimum classification interval functions between the sample sets can be obtained by Eq. (2.1), so that the sample sets are correctly classified and the recreational route can be identified.

3. Research cases

3.1. **Data Sources.** The public transportation network in Xiamen City includes 6738.09 km, and in October 2018, the rate of travel by public transportation was 76.3%. [15] This research is based on Xiamen intelligent transportation card data from October 2018, 397 lines, and vehicle GPS data from October 1, 2018. During the National Day golden week, there are more than 4.31 million credit card data, and there are more than 1.63 million card data records on weekdays, and more than 1.53 million on weekends. Tables 1-4 display the dataset used in this research. The procurement of OD data, which has been finalized for various studies, was accomplished through the evaluation of smart card and GPS data. [4]

(1) Bus GPS arrival data

The complete GPS arrival data of a certain bus schedule includes fields such as route, direction, station number, station name, license plate number, station longitude (all longitude in this article are east longitude) and station latitude (all latitude in this article are north latitude), station number, arrival date, shift, arrival time, etc. (Table 2).

TABLE 1. Sample data

Line number	Fare (yuan)	First bus time (second)	Last bus time (second)	Minimum departures	Maximum departure	Number of shopping centers	Sights Number	Number of external traffic	Number of stations in the shopping mall buffer zone	Number of sites in the scenic spot buffer	Number of stops in the external traffic buffer zone	Maximum number of stations in the buffer
63	2	18133	85254	12	20	9	8	5	36	46	15	46
231	2	21278	85439	10	25	6	12	2	29	64	5	64
162	1	21777	84099	10	20	8	8	3	39	47	14	47
109	1	22114	82937	12	20	5	10	1	21	73	4	73
233	1	21455	82915	6	12	5	9	1	19	67	2	67
65	2	19232	78079	7	15	7	3	5	19	5	16	19
99	1	20709	81426	10	15	9	1	3	37	12	18	37
392	1	20951	83505	15	20	6	4	3	15	29	9	29
402	1	21259	84001	20	30	6	4	3	22	20	10	22
229	1	22893	82583	10	20	4	7	2	21	39	6	39
389	1	21037	85294	8	15	8	2	3	38	7	14	38
525	1	20856	80363	13	20	7	4	2	32	23	12	32
355	2	20365	86169	30	30	7	6	0	22	37	0	37
119	1	19910	86290	6	12	6	4	2	33	15	6	33
130	1	20819	84291	8	12	7	4	1	23	25	2	25
103	1	20513	84043	14	20	3	7	2	15	38	5	38
261	1	22821	81784	8	15	6	4	2	27	25	8	27

NOTE: THE DATA LISTED IN TABLE 1 IS PART OF THE DATA. THE DATA COMES FROM THE GPS DATA OF THE VEHICLE ON OCTOBER 1, 2018 AND THE DATA OF THE INTELLIGENT TRANSPORTATION CARD OF OCTOBER 2018. THE LAST BUS TIME IS THE LAST TIME THE LOCAL BUS SWIPED THE CARD ON THE DAY, IN SECONDS.

TABLE 2. Bus GPS data

Line no.	Direction	Stop index	Stop name	Bus license plate no.	Longitude	Latitude	Date	Vehicle operation shift	Timestamp
9*5	up	1	A	MD001**	118.056	24.618	2018.8.8	12	8:00:15
9*5	up	2	B	MD001**	118.057	24.614	2018.8.8	12	8:06:45
9*5	up	3	C	MD001**	118.060	24.612	2018.8.8	12	8:08:55
9*5	up	4	D	MD001**	118.069	24.601	2018.8.8	12	8:15:10
9*5	up	5	E	MD001**	118.093	24.604	2018.8.8	12	8:20:27

(2) Static bus route station information

The static bus route station information includes fields such as route, direction, station number, station name, station longitude, and station latitude (as shown in Table 3), and the static bus route station information of all operating vehicles on the same route and direction is consistent.

(3) Public transportation IC card data

The original public transportation IC card data may have values that exceed the range, so preprocessing processes such as data cleaning are required to obtain the public transportation IC card data used in this chapter, which includes fields such

TABLE 3. Static bus stop information

Line no.	Direction	Stop index	Stop name	Longitude	Latitude
9*5	up	1	A	118.056	24.618
9*5	up	2	B	118.057	24.614
9*5	up	3	C	118.060	24.612
9*5	up	4	D	118.069	24.601
9*5	up	5	E	118.093	24.604

as card number, card type, card swiping date, card swiping time, route, direction, license plate number, and cost (as shown in Table 4).

TABLE 4. Raw smart card records

ID	Card type	Boarding date	Transaction time	Line no.	Direction	Bus license plate no.	Cost
801**350	Common	2018.8.8	8:06:48	9*5	up	MD001**	0.8
801**170	Common	2018.8.8	8:06:51	9*5	up	MD001**	0.8
801**630	Common	2018.8.8	8:06:54	9*5	up	MD001**	0.8

Selecting tourist attractions, shopping centers, bus stations, sports stations, and airports in Xiamen as recreational areas (Table 5), the bus stop routes and passenger travel chains within a buffer zone of 1000 meters are the total dataset. It is known that Xiamen's tourism dedicated line and ordinary bus routes have significant differences in ticket prices, departure times, departure schedules, and the number of tourist attractions passed through. Therefore, this article selects three datasets: "Bus GPS arrival data", "Bus static route station information", and "Bus IC card data": ticket prices, route numbers, first and last bus times, minimum and maximum departure schedules, as well as scenic spots and shopping centers Six characteristic values of the number of bus stops within the 1000 meter buffer zone of the station (Table 1) were established. A machine learning model dataset was used to identify recreational attribute routes. The dataset was divided into training and testing sets, with the training set used to train the SVM model and the testing set used to test the accuracy (Accuracy) and AUC (Area Under Curve) of the model.

3.2. Route Identification Process.

Step 1: Feature value extraction. This article selected six feature values from three datasets: "Bus GPS Arrival Data", "Bus Static Route Station Information", and "Bus IC Card Data", including ticket price, first and last bus times, minimum and maximum departure times, as well as the number of bus stops within a 1000 meter buffer zone in scenic areas, shopping centers, and stations.

Step 2: Data preprocessing and model construction. We have normalized the features, such as ticket price and time, which have significantly different values, and have trained the model multiple times. This study selected 30% of the total dataset as the training set and 70% as the testing set. Model training utilizes network search and five fold cross validation (CV) on the training set to find the optimal parameters and prevent overfitting of the model.

Table 5. Recreation destinations

Types	Destination
Sights	Gulangyu Island, Zengcuo'an, Nanputuo Temple, Wanshi Botanical Garden, Xiamen Fangte Dream Kingdom, Hulishan Fort, Xiamen University, Tongan Film and Television City, Guanyin Mountain, Jimei Village, Bailuzhou Park, Xiamen Museum, Wuyuan Bay Wetland Park, Hongshan Park, Xiamen Railway Cultural Park, Jinbang Park, Huwei Mountain Park, Huli Park, Xianyue Park, Jiageng Park
Shopping center	Zhongshan Road Shopping Center, China Town, SM City Plaza, Huli Wanda Plaza, Wuyuan Bay Leduhui Shopping Center, Robinson Shopping Plaza, Aluohai City Plaza, Panji Brand Center, China Resources Vientiane City, Baolong One City
External traffic	Xiamen Station, Xiamen North Station, Gaoqi T3 Terminal, Gaoqi T4 Terminal, Wucun Bus Station, Fanghu Passenger Transport Center, Gaoqi (Zhongpu) Bus Station, Jimei Coach Station, Xinglin Coach Station, Tongan Coach station

Step 3: Analysis of model evaluation indicators. Accuracy and AUC (Area Under Curve) are used as evaluation metrics to measure the overall classification accuracy of the model. AUC provides a more comprehensive evaluation, taking into account the performance of the model at various thresholds.

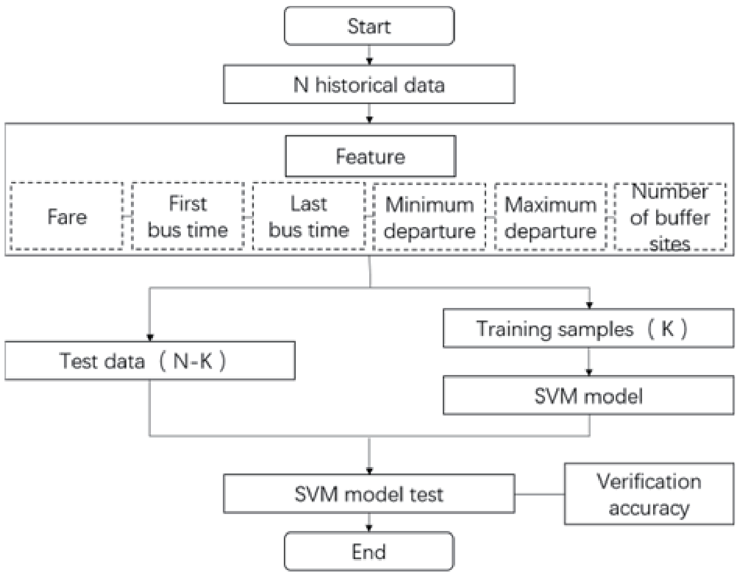

Figure 2. SVM model algorithm implementation steps

3.3. **Model Validation.** In our scenario, we will utilize Xiamen bus lines to evaluate the SVM model and compare it to two traditional machine learning algorithms, KNN and decision trees (DTs). The training set has undergone five-fold cross validation (CV) and network search to minimize over fitting and determine optimal parameters. The major hyper parameter descriptions, search range, and optimized values employed by the three methods are listed Table 6.

Table 6. Hyperparameters for SVM, KNN & DTs

Method	Hyperparameter	Description	Search Range	Optimized Value
SVM	C	Regularization parameter	0.1-200	1
	kernel	Type of kernel function	linear, poly, rbf, sigmoid	poly
	degree	Degree of the polynomial kernel function	1-5	3
	gamma	Kernel coefficient	0.001-0.2	0.08
	coef0	Independent term in kernel function	0-5	3
KNN	n_neighbors	Number of nearest neighbors	1-20	3
	weights	Weight function used in prediction	uniform, distance	distance
	algorithm	Algorithm used to compute the nearest neighbors	ball_tree, kd_tree, brute	kd_tree
DTs	criterion	Function to measure the quality of a split	gini, entropy, log_loss	gini
	splitter	Strategy used to choose the split at each node	best, random	random
	max_depth	Maximum depth of the tree	1-20	10
	min_samples_split	Minimum number of samples required to split an internal node	1-20	5
	min_samples_leaf	Minimum number of samples required to be at a leaf node	1-20	1

This research established the proportion of the number of training sets to the total estimated diverse data volume β in order to thoroughly investigate the parameter impacts under various proportional training sets. However, in order to avoid overfitting and locate the best parameters for the model, the training set underwent fivefold cross validation (CV) and network search. The accuracy and AUC were confirmed throughout the line classification procedure, and the results are shown in Table (7-8). The AUC is 96.4% and the model accuracy is 96.1%. In previous research on tourism route classification, accuracy and indicators were rarely used. Secondly, this study did not use survey questionnaires, which is more efficient and advanced than traditional sampling surveys.

Table 7. Accuracy results under different methods

β	0.15	0.2	0.25	0.3	0.35	0.4
SVM	0.923	0.944	0.9545	0.964	0.903	0.885
KNN	0.555	0.64	0.883	0.875	0.762	0.702
DTs	0.495	0.713	0.773	0.757	0.754	0.769

The comparative findings indicate that the SVM model outperforms the other two approaches in terms of accuracy, with the SVM model showing 96.1%, AUC of 96.4%, KNN model accuracy of 87.5%, AUC of 92.1%, and DTs model accuracy of 75.7%, AUC of 75.9%. The SVM method is better than both the KNN and DT methods. The major emphasis of support vector machines (SVM) is on samples placed at the decision border, or support vectors. This allows SVM to handle situations with dimensions larger than the number of samples well and perform well

TABLE 8. AUC results under different methods

β	0.15	0.2	0.25	0.3	0.35	0.4
SVM	0.916	0.95	0.937	0.961	0.904	0.892
KNN	0.606	0.638	0.921	0.918	0.926	0.889
DTs	0.897	0.696	0.759	0.758	0.778	0.794

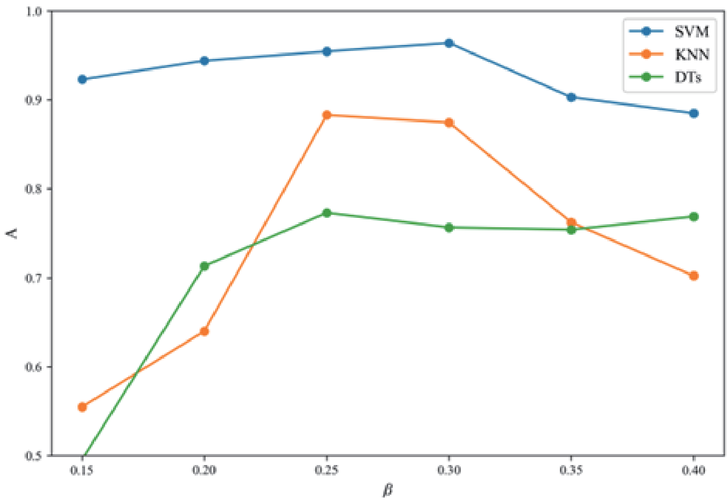

FIGURE 3. The variation of Accuracy with β under different methods

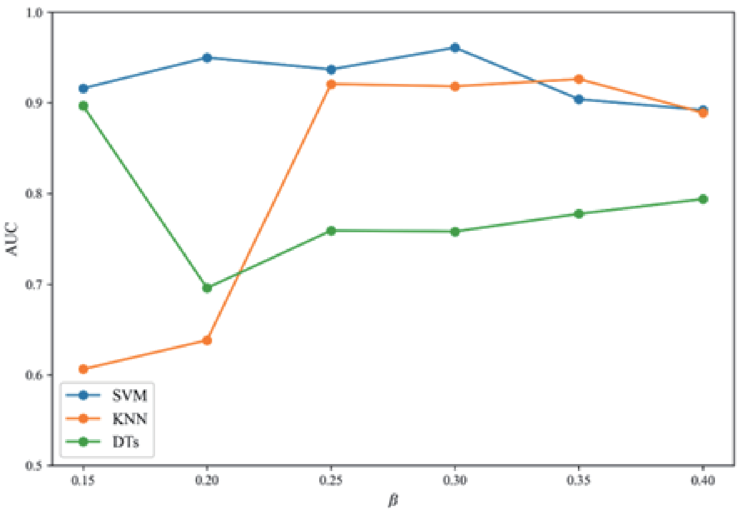

FIGURE 4. The variation of AUC with β under different methods

on small sample datasets. The K value in the KNN algorithm represents the number of neighbors selected, and it is necessary to choose the appropriate K value based on the specific problem. The K value may lead to underfitting or overfitting problems.

Decision trees are very sensitive to small changes in input data, and minor variations in training data could result in completely different tree structures, which can lead to unstable prediction results.

4. Experimental results

This study uses Python language and data processing is based on the machine learning tool library Scikit learn. By solving the SVM model algorithm, 159 recreational lines that basically coincide with recreational space of Xiamen were finally identified, accounting for 40% of bus lines in Xiamen. The solution results are shown in Figure 5.

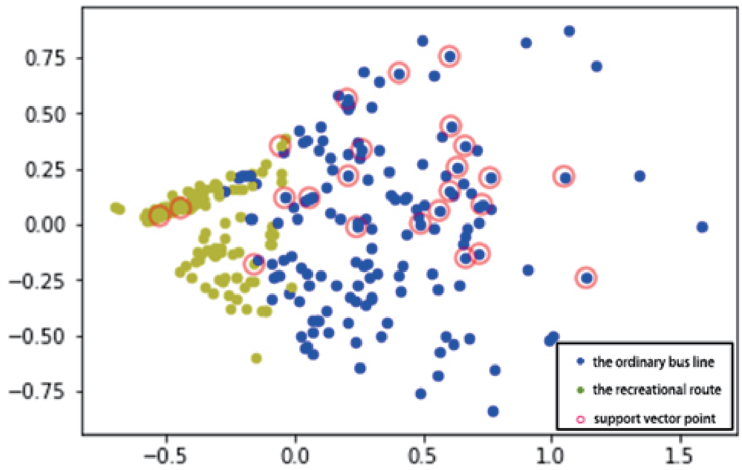

Figure 5. SVM line classification result chart

The set of recreation routes identified in this paper is basically consistent with recreational recreation space of Xiamen, and can reflect the results of this experiment and the actual space. The spatial distribution of the collection of recreational lines in Xiamen is shown in Figure 6 and 7.

5. Research conclusion and discussion

A support vector machine model was developed using the six characteristic variables that this study solved: ticket price, establishing bus time, final bus time, minimum departure frequency, maximum departure frequency, and number of buffer zone stations. Machine learning methods were used to classify and differentiate leisure routes in Xiamen, revealing 159 recreational pathways. The model's validity was confirmed, and big data mining technology was used to identify recreational attribute routes using three datasets: Bus IC card data, Bus static route station information, and Bus GPS arrival data. Travel requirements and characteristics of recreational groups were also considered. In addition to offering a solid empirical foundation for the design of routes for recreational public transportation, this approach offers theoretical backing for additional investigations like urban functional space optimization and transportation network optimization. The implementation

FIGURE 6. Spatial distribution of recreational lines in Xiamen

of a route recognition model has the potential to enhance the efficiency of the urban transportation network and further optimize functional spatial layout of the city.

Three machine learning algorithms, SVM, KNN, and DTs, were used to categorize recreation routes on metropolitan public transit. Further research on tourist leisure activity patterns and traits, focusing on recreational transportation routes, could provide more precise findings. The advancement of big data mining technology allows for a more intuitive understanding of travel trends. Urban recreational public transportation, a component of urban tourism, has potential for growth. Providing services to recreation groups can enhance public transit quality and contribute to urban tourism growth. Improving and making the most of public transportation is essential for urban recreation. The research explores alternative classification methods beyond the support vector machine algorithm, including decision trees, random forests, neural networks, clustering analysis, and association rule mining for line identification. These methods could enhance line classification and identification, providing new insights and potential solutions for urban transportation planning and public transportation service optimization.

The location of this study is Xiamen, China. As a famous tourist city, Xiamen's public transportation is closely related to tourism development. The results of this study aim to provide reference for tourism cities to more effectively identify and plan tourism public transportation routes, especially to optimize the design of tourism dedicated lines based on existing public transportation networks. The classification

Figure 7. Spatial distribution of lines in Xiamen

method proposed in this study can not only help Xiamen further improve the convenience and efficiency of tourism transportation, but also has broad applicability and can be promoted to other tourist cities, providing innovative optimization paths and reference solutions for their public transportation systems.

References

[1] M. Addoun, M. Chachour, and M. S. Zerouala, *A strategy for the design of a cultural tourist route. Case study of the city of Tlemcen in Algeria*, PASOS Revista de Turismo y Patrimonio Cultural **21** (2023), 743–765.

[2] M. Amrani, *Deep convolutional multi-informative metric correlation analysis with bottleneck attention module for face recognition in the wild*, Multimedia Tools and Applications **83** (2024), 62459–62487.

[3] C. Cortes and V. Vapnik, *Support-Vector Networks*, Machine Learning **20** (1995), 273–297.

[4] Z. Cui, C. Wang, Y. Gao, D. Yang, W. Wei, J. Chen and T. He, *Alighting stop determination of unlinked trips based on a two-layer stacking framework*, Mathematical Problems in Engineering **2021** (2021): 6464980.

[5] J. B. Duarte-Duarte, L. H. Talero-Sarmiento and D. C. Rodríguez-Padilla, *Methodological proposal for the identification of tourist routes in a particular region through clustering techniques*. Heliyon **7** (2021): e06655.

[6] R. Gholami and N. Fakhari, *Support vector machine: principles, parameters, and applications*, in: Handbook of Neural Computation, P. Samui, S. Sekhar, V. E. Balas (eds), Academic Press, Elsevier, 2017, pp. 515–535.

[7] J. Guo, *The travel characteristics analysis of passengers based on the bus data*, M.S. thesis, Dept. Electron. CHS., SCUT., Guangzhou, CHN, 2016.

[8] H. Han, X. Yu and Y. Long, *Identifying urban functional zones using bus smart card data and points of interest in Beijing*, City Planning Review **40** (2016), 52–60.

[9] J. Han, *Application of SVM model to environmental resource analysis in tourism development*, J. Phys.: Conf. Ser. **1629** (2020): 012007.

[10] W. A. Khan, M. Masoud, A. E. Eltoukhy and M. Ullah, *Stacked encoded cascade error feedback deep extreme learning machine network for manufacturing order completion time*, Journal of Intelligent Manufacturing **36** (2024), 1313–1339.

[11] Z. Li, Z. Chen and Z. Qin, *Support vector machine in forecasting tourism demand*, Economic Geography **30** (2010), 2122–2126.

[12] S. Liu, R. Liu and N. Tan, *A spatial Improved-kNN-based flood inundation risk framework for urban tourism under two rainfall scenarios*, Sustainability **13** (2021): 2859.

[13] Y. Long, M. Mao, Q. Mao, Z. Shen and Y. Zhang, *Fine-Scale urban modeling and its opportunities in the 'Big Data' era: methods, data, and empirical studies*, Human Geography **29** (2014), 7–13.

[14] Y. Long, Y. Zhang and C. Cui, *Identifying commuting pattern of Beijing using bus smart card data*, Acta Geographica Sinica **67** (2012), 1339–1352.

[15] *October 2018 Xiamen City public transport passenger flow characteristics monthly report*, Urban construction and Economic Development Institute of Huaqiao University, Xiamen, FJ, CHN, Rep. DEC. (2018).

[16] N. Pulmamidi, R. Aluvalu and V. U. Maheswari, *Intelligent travel route suggestion system based on pattern of travel and difficulties*, IOP Conf. Ser.: Mater. Sci. Eng. **1042** (2021): 012010.

[17] A. L. Serrano López, S. A. Freire Chaglla, F. E. Espinoza-Figueroa, D. S. Andrade Tenesaca and M. E. Villafuerte Pucha, *Modeling of tourist profiles with decision trees in a world heritage city: The case of cuenca (ecuador)*, Tourism Planning & Development **16** (2019), 473–493.

[18] S. R. Stephan, *The nature of statistical learning theory*, Technometrics **38** (1997): 409.

[19] P. Thiengburanathum P, S. Cang and H. Yu, *A decision tree based recommendation system for tourists*, in: Proceedings of 2015 21st International Conference on Automation and Computing, IEEE, 2015, pp. 1–7.

[20] V. N. Vapnik, *An overview of statistical learning theory*, IEEE Trans. Neural Net Works. **10** (1999), 988–999.

[21] D. Wang, F. Sun, C. Liu and J. Qian, *Impact of low-carbon public transport system on urban recreational spatial: A case study of center city of Suzhou*, Tourism Tribune **32** (2017), 78–90.

[22] Y. Yin, G. Hu and Y. Qiu, *Tourism demand forecast and analysis to Yunnan based on statistical learning theory*, Journal of Yunnan University (Natural Sciences Edition) **26** (2004): 4.

[23] X. Yu, S. Hu, G. Zhu and D. Li, *Research on medium-term prediction of tourist arrivals in scenic areas based on least squares support vector machines*, Tourism Tribune **28** (2013), 75–82.

[24] Z. Zhou, *Machine Learning*, Beijing: Tsinghua University Press, 2016.

Manuscript received February 10, 2024
revised October 13, 2024

X. Ji
College of Tourism, Huaqiao University, Quanzhou 362021, China; Guangzhou Maritime University, Guangzhou 510000, China
E-mail address: `sdqdjxx@163.com`

A. Huang
College of Tourism, Huaqiao University, Quanzhou 362021, China
E-mail address: `17011121003@stu.hqu.edu.cn`