

# Application of support vector machine in trip chaining pattern recognition and analysis of explanatory variable effects

Yang Shuo Deng Wei Cheng Long

(School of Transportation, Southeast University, Nanjing 210096, China)

**Abstract:** In order to improve the accuracy of travel demand forecast and considering the distribution of travel behaviors within time dimension, a trip chaining pattern recognition model was established based on activity purposes by applying three methods: the support vector machine (SVM) model, the radial basis function neural network (RBFNN) model and the multinomial logit (MNL) model. The effect of explanatory factors on trip chaining behaviors and their contribution to model performance were investigated by sensitivity analysis. Results show that the SVM model has a better performance than the RBFNN model and the MNL model due to its higher overall and partial accuracy, indicating its recognition advantage under a small sample size scenario. It is also proved that the SVM model is capable of estimating the effect of multi-category factors on trip chaining behaviors more accurately. The different contribution of explanatory factors to trip chaining pattern recognition reflects the importance of refining trip chaining patterns and exploring factors that are specific to each pattern. It is shown that the SVM technology in travel demand forecast modeling and analysis of explanatory variable effects is practical.

**Key words:** trip chaining patterns; support vector machine; recognition performance; sensitivity analysis

**DOI:** 10.3969/j.issn.1003-7985.2017.01.018

Trip chaining behavior is the fact that residents always connect several activities with different purposes that need to be done in a day within temporal and spatial constraints. Contrary to individual single trip behavior, trip chaining behavior contains a series of space, time and activity type information. Therefore, with the development of the activity-based theory, there is much research that emphasizes an analysis of which factors have effects on trip chaining behaviors. For example, both Xianyu<sup>[1]</sup> and Li et al.<sup>[2]</sup> used the co-evolutionary logit model to capture the interrelationship between the travel mode choice and

trip chaining behavior, and the model results show that activity decision or trip chaining behavior is made before mode choice in most cases. Meanwhile, household characteristics and personal attributes are also found significantly influencing trip chaining behavior, such as household members' work status<sup>[3]</sup>, work schedule<sup>[4]</sup>, gender<sup>[5]</sup>, the presence of children<sup>[6]</sup>, etc. In addition, the built environment where people live and work is another important factor, and its effects on trip chaining behaviors have found their way into the diverse concepts of planning, design and policies<sup>[7-9]</sup>, such as a compact city, transit-oriented development and mixed land use.

However, as a matter of fact, all aspects of factors have influences on trip chaining behavior collaboratively, rather than separately, which means all these factors need to be carefully considered when studying trip chaining behavior. Furthermore, although works on the relationship between factors and chaining behavior have obtained rich results, there are few studies focusing on developing effective recognition models for trip chaining patterns. Given these circumstances, the primary objective of this paper is to experiment with different technologies to recognize residents' trip chaining behavior, particularly to investigate the potential capability of using the support vector machine (SVM) algorithm, which is new data-mining technology. Specifically, the study includes the following tasks: comparing the recognition performance of statistical and machine learning methods; evaluating the impact of those possible explanatory variables on trip chaining choice behavior and comparing the results between different methods and putting forward some predictor selection improvements for forecast models' accuracy.

## 1 Methodology

### 1.1 The SVM model

The SVM model is an advanced technology of data mining based on the statistical learning theory and the structural risk minimization principle<sup>[10]</sup>, which is capable of approximating linear and nonlinear relationships<sup>[11]</sup>. Through constructing an optimal hyper-plane with kernel functions, the SVM model can change a nonlinear problem into a linear problem and then realizes its regression or classification function. Clearly, trip chaining pattern recognition should be treated as a classification problem in this paper. There are usually four main steps for process-

Received 2016-06-20.

**Biographies:** Yang Shuo (1988—), male, graduate; Deng Wei (corresponding author), male, doctor, professor, dengwei@seu.edu.cn.

**Foundation items:** The Fundamental Research Funds for the Central Universities, the Scientific Innovation Research of College Graduates in Jiangsu Province (No. KYLX\_0177).

**Citation:** Yang Shuo, Deng Wei, Cheng Long. Application of support vector machine in trip chaining pattern recognition and analysis of explanatory variable effects[J]. Journal of Southeast University (English Edition), 2017, 33(1): 106 – 114. DOI: 10.3969/j.issn.1003-7985.2017.01.018.

ing recognition.

1) Sample division: The complete data samples are randomly divided into two subsets, which are a learning subset and a holdout subset.

2) Training model: The trip chaining pattern recognition SVM model is trained based on the learning subset to obtain structural parameters.

3) Recognition: The SVM model uses these structure parameters again but it is fitted by the explanatory variables of the holdout subset. In this way, the trip chaining patterns for each sample (individual) in holdout subset can be recognized.

4) Evaluation: The performance of the SVM model can be measured via checking whether the recognized trip chaining pattern matches the real trip chaining pattern for each sample (individual).

## 1.2 Grid-search and 10-fold cross-validation

Generally, there are structural parameters that need to be calculated for optimizing the SVM performance, for example, the penalty factor that determines the trade-off between the complexity of model and training errors and the parameter in the kernel function. The radial basis kernel function is used in this paper. A combination of grid-search and cross-validation is applied to calibrate these parameters.

10-fold cross-validation is an effective experimental method to estimate generalization error in a pattern recognition problem in the training model. It subdivides the learning set into 10 subsets, and each subset is tested using the SVM model and trained on the remaining nine subsets sequentially, thus ten-time trainings are processed. After averaging the percentages of samples that are correctly recognized in all trainings, the cross-validation accuracy can be obtained.

The grid-search method specifies a plane which is the scope of two-dimensional parameters. Through setting up the search step length of parameters, the plane is partitioned into a number of grids, which represent various pairs of parameters. Each pair of parameters is taken into the cross-validation process, after searching all grids in the plane, the pair with the best cross-validation accuracy, namely the pair of optimal parameters, is determined.

## 1.3 Methods for comparison purpose

In order to verify whether the SVM model has the advantage of recognizing trip chaining patterns, another two methods are selected for comparison purpose, which are the multinomial logit (MNL) model and a typical three-layer framework of the radial basis function neural network (RBFNN) model. Since the MNL model produces probability of each trip chaining pattern, the pattern with the highest probability should be selected, and then the

recognition function of the MNL model can be calculated.

## 2 Data Source

The original data for this paper was collected from the 2013 Nanjing resident trip survey, which was conducted in the main city area of Nanjing consisting of 495 traffic analysis zones (TAZs) according to the street and administrative divisions. Above six-year-old residents were asked to provide individual characteristics, household characteristics, and all one-day trip information on October 30th, 2013. Initially, a total of 2 000 household attributes and 5 932 residents' personal characteristics and travel information were collected. Samples with missing information and logical issues was excluded. Finally, 1 701 households', 3 118 residents' information was selected in this study.

According to Ref. [2], the trip chaining behavior is usually defined as a sequence of trips that occurs on a complete day which starts and ends at home. Thus, three major trip chaining patterns and five detailed trip chaining patterns are given in Tab. 1. They are described as follows:

- HW + WH denotes that there are only subsistence activities (such as working, school, and business) within a day. It may contain one commuting activity stop, HWH, or more than one stops, HW + H.

- HO + OH denotes that there are only non-subsistence activities (such as shopping, entertainment, visiting friends, and picking up/dropping off) within a day. It may contain one non-commuting stop, HOH, or more than one stops, HO + H.

- HW + OH denotes that this chaining pattern is a combination of a subsistence chain and a non-subsistence chain with at least two stops.

**Tab. 1** Description of trip chaining patterns

Major Patterns description	Detail patterns description	Sample size	Percent/%
HW + WH	HWH	1 862	59.7
	HW + H	148	4.8
HO + OH	HOH	531	17.0
	HO + H	253	8.1
HW + OH	HW + OH	324	10.4

Tab. 2 reports all explanatory variables for model specification, including three aspects: trip information, personal and household characteristics and land use characteristics. Of the trip information, travel mode choice is divided into two types, which are single-mode and multi-modes. The former denotes that a person finishes all trips only by one mode, while the latter denotes two or more modes. Transfer times sum the count of transfer in each trip. Road network saturation reflects the average traffic saturation of traffic analysis zone where residents live. Land use characteristics show some information of residents' dwelling environments.

**Tab. 2** Explanatory variables for model development

Factor category	Variables	Description	Sample size	Percent/%
Trip information	Average duration/min	Numeric ( continuous)	31. 6 ( Mean)	
	Transfer times	Numeric ( count)	0. 05 ( Mean)	
	Road network saturation	Numeric ( continuous)	0. 67 ( Mean)	
	Crossing group travel	No	1 935	62. 1
		Yes	1 183	37. 9
	Mode choice	Single-mode	2 808	90. 1
Multi-modes		310	9. 9	
Personal attributes	Gender	Male	1 539	49. 4
		Female	1 579	50. 6
	Job	Employee	2 036	65. 3
		Self-employed	295	9. 5
		Retired	787	25. 2
	Age	15 to 29 years old	510	16. 4
		30 to 49 years old	1 468	47. 1
		≥50years old	1 140	36. 6
	Public card possession	Yes	2 626	84. 2
		No	492	15. 8
	Car license possession	Yes	1 452	46. 6
		No	1 666	53. 4
	Education	High school or lower	1 517	48. 7
		college or higher	1 601	51. 3
Household characteristics	Head of household	Yes	1 186	38. 0
		No	1 932	62. 0
	Family size	≤3	2 373	76. 1
		> 3	745	23. 9
	Child in household	No	2 632	84. 4
		Yes	486	15. 6
	Cars ownership	No	1 478	47. 4
		Yes	1 640	52. 6
	Bikes ownership	No	902	28. 9
		Yes	2 216	71. 1
	Ebikes ownership	No	929	29. 8
		Yes	2 189	70. 2
	Income	< 10 <sup>5</sup> yuan/year	1 732	55. 5
		≥10 <sup>5</sup> yuan/year	1 386	44. 5
Land use characteristics	Population density/( persons · km <sup>-2</sup> )	Numeric ( continuous)	15. 3 ( Mean)	
	Employment density/( jobs · km <sup>-2</sup> )	Numeric ( continuous)	7. 4 ( Mean)	
	Living in center area	No	911	29. 2
		Yes	2 207	70. 8

3 Results

In this paper, the SVM model, as well as the MNL model and the RBFNN model, was experimented on ten times with different learning and holdout sets with the ratio of 4: 1 randomly, in order to reduce the bias when producing the random separation of the whole dataset. The SVM toolbox developed by Chang et al. <sup>[12]</sup> was used to specify the SVM model whose parameters were optimized by the grid-search approach. The MNL model and the RBFNN model were fitted using the SPSS software packages.

These recognition models’ performance is investigated from two sides: overall accuracy and partial accuracy. The former is defined as the proportion of all-class that

recognized actually in the whole holdout sets, while the latter focuses on the proportion of each class that recognized actual accounts for every class of patterns in the holdout sets.

3.1 Accuracy assessment and comparison of recognition performance

The partial recognition performance is represented by the confusion matrix, which contains two evaluation measures: the precision *P* for one trip chaining pattern *X* denotes the proportion of recognized trip chaining pattern *X* that correctly reflects trip chaining pattern *X*; the recall *R* for one trip chaining *X* denotes the proportion of real trip chaining pattern *X* that correctly recognizes the trip chaining pattern *X*. For comparison purposes, all ten ex-

perimental results produced by the three models were summarized into one confusion matrix as shown in Tab. 3 and Tab. 4.

For major trip chaining pattern recognition, it is clear that, of the recall percentage, all models have good performance in identifying the HO + OH trip chaining pattern. Compared with the MNL model and the RBFNN model, the SVM model performs much better in recognizing the HW + OH trip chaining pattern. Of the precision, all models have more than 75% percentage, except that the precision of the MNL model for HO + OH is less than 30%. For detailed five-category trip chaining patterns, the SVM model, the MNL model and the RBFNN model experience great difficulty in identifying the trip chaining pattern of HW + H. However, the recognition

performance of the SVM model is higher in terms of precision and recall percentages than those of the other two models.

Although Tabs. 3 and 4 provide much information about the performance of different models, in order to determine their effectiveness, we compare all the models by using  $F$ -measure<sup>[13]</sup>, which takes both the precision and recall percentages into consideration and it is computed as

$$F = \frac{2PR}{P + R} \quad (1)$$

The  $F$ -measure values in Tabs. 3 and 4 indicate that the SVM model performs better than the MNL model and the RBFNN model in terms of different trip chaining divisions and patterns.

**Tab. 3** Confusion matrix for the three models developed for recognizing the three major trip chaining patterns

Model	Actual trip chaining pattern	Recognized trip chaining pattern			$R/\%$	$F$
		HW + WH	HO + OH	HW + OH		
SVM	HW + WH	3825	114	56	95.74	0.93
	HO + OH	123	1473	8	91.83	0.91
	HW + OH	264	38	329	52.14	0.64
	$P/\%$	90.81	90.65	83.72		
MNL	HW + WH	617	3371	7	15.44	0.26
	HO + OH	28	1575	1	98.19	0.45
	HW + OH	31	497	103	16.32	0.28
	$P/\%$	91.27	28.94	92.79		
RBFNN	HW + WH	3767	193	35	94.29	0.90
	HO + OH	186	1402	16	87.41	0.86
	HW + OH	419	43	169	26.78	0.40
	$P/\%$	86.16	85.59	78.82		

**Tab. 4** Confusion matrix for the three models developed for recognizing detailed three trip chaining patterns

Model	Actual trip chaining pattern	Recognized trip chaining pattern					$R/\%$	$F$
		HWH	HOH	HW + H	HO + H	HW + OH		
SVM	HWH	3527	102	1	1	51	95.79	0.89
	HOH	100	941	0	30	0	87.86	0.75
	HW + H	295	13	0	0	5	0	
	HO + H	19	366	0	139	9	26.08	0.39
	HW + OH	269	23	2	13	324	51.35	0.64
	$P/\%$	83.78	65.12	0	75.96	83.29		
MNL	HWH	695	2726	0	238	23	18.88	0.31
	HOH	28	947	0	96	0	88.42	0.34
	HW + H	16	248	0	47	2	0	
	HO + H	1	336	0	192	0	36.02	0.28
	HW + OH	38	219	0	271	103	16.32	0.27
	$P/\%$	89.33	21.16		22.75	78.03		
RBFNN	HWH	3522	125	0	1	34	95.65	0.87
	HOH	160	902	0	6	3	84.22	0.69
	HW + H	281	28	0	0	4	0	
	HO + H	55	442	0	14	22	2.63	0.05
	HW + OH	435	32	0	1	163	25.83	0.38
	$P/\%$	79.09	58.99		63.64	72.12		

In addition to partial accuracy, the average overall accuracy of all experiments is also compared. For major trip chaining patterns, the average overall accuracy of the SVM model, the MNL model and the RBFNN model are 90.32% (0.010 6), 36.84% (0.024 8), 85.68%

(0.016 1), respectively, which in parentheses are standard errors. For detailed trip chaining patterns, the average overall accuracy of the SVM model is 79.15% (0.016 5), the MNL model's is 31.09% (0.033 4), and the RBFNN model's is 73.85% (0.019 0). The result clear-

ly shows that the SVM model outperforms another two models no matter how the trip chaining patterns are divided and the lower standard deviation demonstrates that the outcomes of the SVM model are more stable. In conclusion, it is effective to use the support vector machine learning technique to produce acceptable and accurate outcomes in trip chaining pattern recognition.

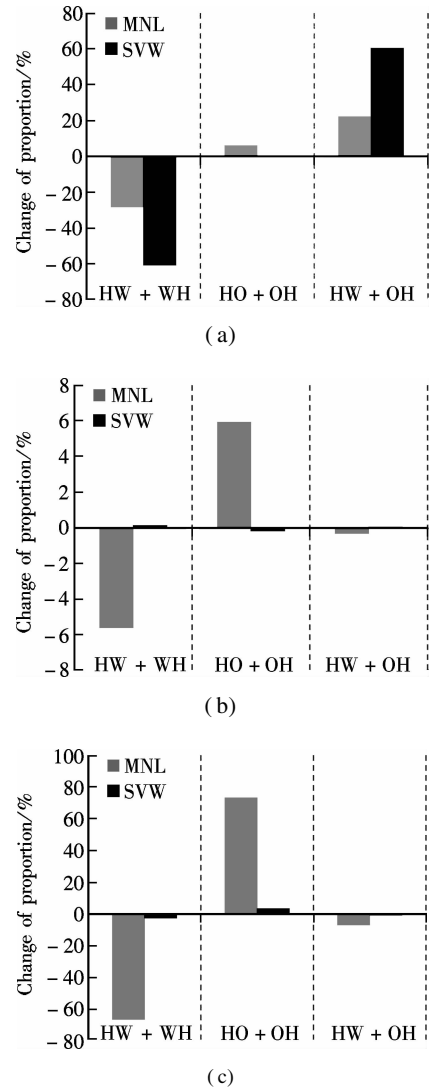
### 3.2 Analysis of effects of explanatory variables

Although the SVM model performs well in terms of partial accuracy and overall accuracy, impact evaluation of explanatory variables is a common problem in machine learning methods, such as the SVM model and the neural network model, due to their “black-box” mechanism. However, inspired by Refs. [10, 14], the one-dimensional sensitivity analysis of the SVM model is conducted in this study to investigate the relationship between trip chaining behavior and various explanatory variables. Each explanatory variable fitted in the SVM model is perturbed by increasing one unit change with other variables remaining unchanged, and then the proportion change of trip chaining patterns is estimated by comparing the SVM’s recognition result before and after the perturbation. Furthermore, because there are 21 explanatory variables, the sensitivity degrees caused by all the variables are ranked from the maximum variation to the minimum variation. Sensitivity degree is defined as the sum of absolute value of probability changes of all-category trip chaining patterns. Each variable’s sensitivity degree is then used to calculate its contribution that accounts for the whole SVM model sensitivity degree gross, and only these ranked variables whose accumulative contribution proportion is above 95% are concerned as illustration. Meanwhile, the marginal effect of each input variable on each trip chaining pattern was also estimated using the MNL model, for comparison between machine learning and statistical models.

#### 3.2.1 Sensitivity analysis result of the SVM model for three major trip chaining patterns

Sensitivity analysis results of the SVM and the MNL models for three major trip chaining patterns are presented in Fig. 1. Only two explanatory variables, which are mode choice and job, have the accumulative sensitivity proportion occupancy above 95% (96.97%). As shown in Fig. 1, the SVM model and the MNL model produce a similar results for the variable of mode choice, however, they produce opposite results for the variable of the job. In order to further explore and explain underlying reasons, the proportion of three chaining patterns grouped by different mode choice and jobs based on the complete data are plotted respectively in Fig. 2.

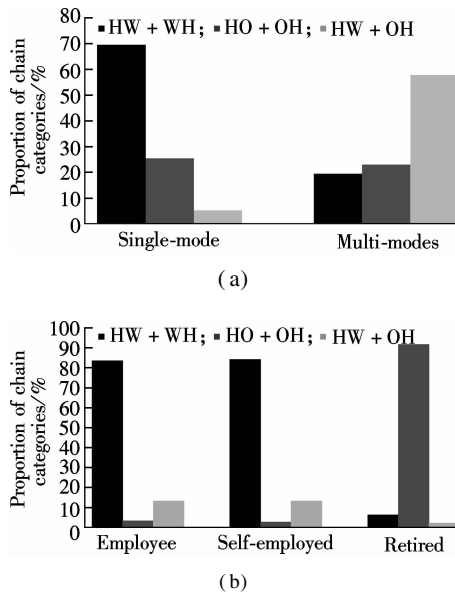
For mode choice, if an individual changes his/her mode type from single-mode to multi-modes, he/she seldom makes the complex subsistence-related trip chain of



**Fig. 1** Comparison of the SVM and the MNL models regarding the impacts of mode choice and job on major trip chaining patterns. (a) Mode choice single to multiple; (b) Job employee to self-employed; (c) Job employee to retired

HW + WH (the proportion of HW + HW chaining pattern decrease from 69.4% to 19.4%) but tends to make a mixed trip chain of HW + OH (the proportion of HW + OH chaining pattern increase from 5.2% to 57.7%). Both the SVM and the MNL models captured the significant changes. However, according to Fig. 2, people who use multi-modes will not like to make the trip chain of HO + OH (the proportion of HO + OH chaining pattern decrease slightly by 2.5% when comparing people who use multi-modes with who use single-mode), two models cannot reflect the subtle probability change of the HO + OH pattern, and even the MNL model has positive change which is not accurate.

For the job, it is not difficult to find that the MNL model keeps the similar trend no matter how an individual’s job changes, and clearly the sensitivity result of the MNL model is much more significant than the SVM’s. However, the sensitivity result of the SVM model



**Fig. 2** Proportion of three chain categories. (a) Grouped by different mode choices; (b) Grouped by different jobs

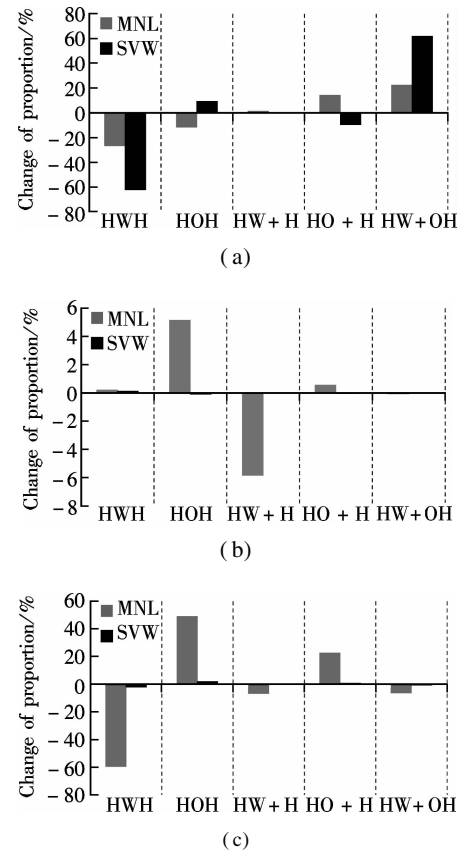
is only consistent with the MNL's when an individual's job changes from employee to being retired. If someone's job changes from employee to self-employed, two models show the opposite result. From Fig. 2, it seems that the SVM offers more accurate sensitivity. When comparing an individual whose job is self-employed with the one whose job is employee, the former tends to make a little more trip chaining behavior of HW + WH (84.1% vs. 83.4%), which corresponds to the positive change in proportion of HW + WH as displayed in Fig. 1, but makes a little less trip chaining behavior of HO + OH (2.7% vs. 3.3%), which corresponds to the negative change of proportion of HO + OH as is also displayed in Fig. 1. Therefore, the SVM model has the capability of capturing the interrelationship among multi-category explanatory variables particularly when the difference between each category is imperceptible.

### 3.2.2 Sensitivity analysis result of the SVM model for five detailed trip chaining patterns

The SVM model for five detailed trip chaining patterns was also analyzed by sensitivity calculation. There are six explanatory variables, mode choice, job, average travel duration, family size, transfer times and population density, which make the accumulative sensitivity occupancy extend beyond 95% (95.54%). Also, the variables of mode choice and job that are the two most sensitive variables are regarded as examples for interpreting the results.

By exploring the description in Fig. 3 and Fig. 4, what should be noted is that generally sensitivity estimated by the SVM model is only suitable for analyzing the impact of explanatory variables on trip chaining patterns of HWH and HW + OH as the number of trip chaining pattern classification increases. However, the SVM model can still produce more accurate sensitivity for the patterns of HOH

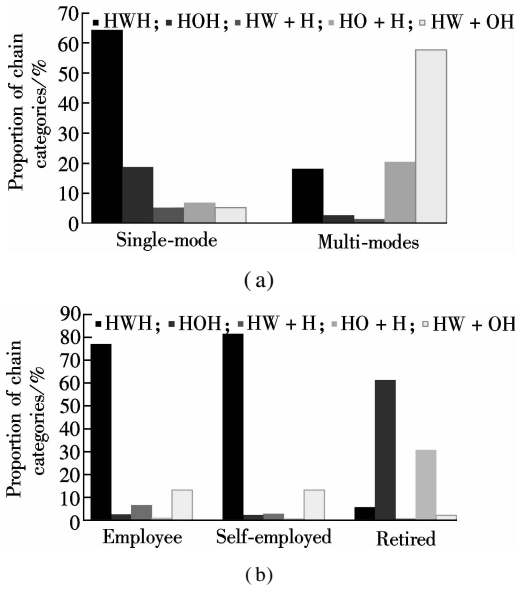
and HO + H when analyzing multi-category explanatory variables such as the variable of job. The impact of variables, mode choice and job types, on trip chaining pattern of HWH and HW + OH, conducted by the SVM model is consistent with the estimation of the MNL model. However, the SVM model fails when investigating the impact of explanatory variables on the trip chaining pattern of HW + H; that is, no matter how much the variables change, the proportion of HW + H in the SVM model recognition output will not change, indicating that none of the explanatory variables used in this paper can influence this type of trip chaining behavior. This unexpected result may be due to the minimum sample size of HW + H in the complete data and the fact that the SVM model suffers from a multi-class classification problem. Consequently, it can be concluded that when conducting the analysis on explanatory variable impacts, various methods need be applied not only for comparison but also for obtaining more reliable results.



**Fig. 3** Comparison of the SVM and the MNL models regarding the impacts of variables of mode choice and job on detailed trip chaining patterns. (a) Mode choice single to multiple. (b) Job employee to self-employed. (c) Job employee to being retired

### 3.3 Analysis of explanatory variables' contribution to model recognition performance

During the sensitivity analysis, the percentage change of the SVM models' recognition accuracy also based on the



**Fig. 4** Proportion of five chain categories. (a) Grouped by different mode choices; (b) Grouped by different jobs

complete dataset is recorded, as shown in Tab.5 and Tab.6.

For the major trip chaining pattern model, it shows that job and mode choice greatly aid the improvement of the model’s overall accuracy performance, while other variables make little contribution. Of specific trip chaining patterns, only transfer times and job are significantly helpful for HW + WH pattern recognition accuracy. All of the variables improve the HO + OH patterns recognition accuracy, but most of them are much too small. Mode choice makes a great contribution to improving the accuracy of the HW + OH pattern by 49.07%. Conversely, the mode choice variable has a negative effect on HW + WH patterns recognition accuracy. This is consistent with the actual data. For example, most people whose trip chaining patterns are HW + WH and HO + OH always travel by single mode, while more people whose trip chaining pattern is HW + OH take multi-modes, which may cause the model difficulty when recognizing the patterns of HW + WH or HO + OH when one person travels by single mode.

**Tab.5** Predictor contributions to the recognition performance of the SVM model for major trip chaining patterns

Factor category	Variables	Variable importance/%			
		For overall	For HW + WH	For HO + OH	For HW + OH
Trip information	Average duration	0.320 7	0.096 5	0.959 4	1.234 5
	Transfer times	0.673 5	0.942 3	0.576 8	0.308 6
	Road network saturation	0	-0.003 0	0.449 2	0
	Crossing group travel	0.032 1	-0.003 0	0.449 2	0.308 6
	Mode choice	4.329 7	-1.246 9	0.576 8	49.074 1
Personal attributes	Gender	0.064 2	-0.003 0	0.576 8	0.308 6
	Job	5.580 5	5.867 6	4.785 9	6.790 1
	Age	0	0.046 7	0.449 2	-0.308 7
	Public card possession	0.032 1	-0.003 0	0.449 2	0.308 6
	Car license possession	0.096 2	0.046 7	0.576 8	0.308 6
	Education	0.032 1	-0.003 0	0.449 2	0.308 6
Household characteristics	Head of household	0.032 1	0.046 7	0.449 2	0
	Family size	0.064 2	0.046 7	0.449 2	0.308 6
	Child in household	0.096 2	0.096 5	0.449 2	0.308 6
	Cars ownership	0.096 2	0.096 5	0.449 2	0.308 6
	Bikes ownership	0	-0.003 0	0.449 2	0
	Ebikes ownership	0.032 0	-0.003 0	0.321 7	0
Land use characteristics	Income	0.032 1	-0.003 0	0.449 2	0.308 6
	Population density	-0.096 2	-0.102 5	0.321 7	0
	Employment density	0.032 1	-0.003 0	0.449 2	0.308 6
	Living in center Area	0	-0.004 8	0.449 2	0

For the detailed trip chaining pattern model, the recognition performance change becomes much more complex. Similar to the major trip chaining pattern model, job and mode choice still have significant positive effects on overall recognition accuracy. Meanwhile, more variables shows slight contribution to overall recognition accuracy such as average duration, transfer times, education, family size, e-bike ownership and population density. Interestingly, some of these predictors show the opposite impact on different types of trip chaining patterns. For example, the variable of mode choice has a positive impact

on recognizing HW + H and HW + OH patterns, but it affects HWH and HO + H patterns negatively. Besides, more predictors selected in this paper are not able to helpfully recognize the pattern of HW + H, and with the increase of categories being divided, fewer variables demonstrate positive effects on the recognition of the HW + OH pattern, although its absolute sample size is unchanged no matter how the trip chaining patterns are divided.

In summary, two conclusions can be drawn: different explanatory variables may be specific to trip chaining pat-

terns and their sub-division influences the contribution to recognition performance. Therefore, it can be said that trip chaining patterns should be divided extensively and

the explanatory variables' effects must be determined carefully.

**Tab. 6** Predictor contributions to the recognition performance of the SVM model for detailed trip chaining patterns

Factor category	Variables	Variable importance/%				
		For overall	For HWH	For HW + H	For HO + H	For HW + OH
Trip information	Average duration	1.282 9	0.053 7	-1.130 0	16.205 5	1.234 5
	Transfer times	0.859 6	1.396 3	0.188 3	0	0
	Road network saturation	0.032 1	0	0.188 3	0	0
	Crossing group travel	0.096 2	0.053 7	-0.376 7	1.581 0	0
	Mode choice	5.452 2	-1.074 2	4.519 8	-0.395 3	51.543 2
Personal attributes	Gender	0.160 3	0.053 7	0.753 3	0	0
	Job	4.714 5	5.585 3	5.649 7	1.976 2	2.46 91
	Age	0.096 2	0.053 7	0.376 6	0	0
	Public card possession	0.128 3	0.053 7	0.376 6	0.395 2	0
	Car license possession	0.032 1	0.053 7	0.188 3	-0.395 3	0
	Education	0.224 5	0.107 4	0	1.976 2	0
	Head of household	0.096 2	0.053 7	-0.188 3	1.581 0	-0.308 7
Household characteristics	Family Size	0.160 3	0.053 7	-0.188 3	-0.023 8	0
	Child in household	0	0.053 7	5.811 7	0	0
	Cars ownership	0.128 3	0.053 7	-0.376 7	1.581 0	0.308 6
	Bikes ownership	0.160 3	0.107 4	-0.188 3	1.581 0	0
	Ebikes ownership	0.352 8	0.107 4	0.376 6	2.766 7	0
	Income	0.064 1	0	0.1883	0	0.308 6
Land use characteristics	Population density	0.384 8	0.107 4	-0.941 6	5.533 5	0.308 6
	Employment density	0.032 1	0.053 7	-0.188 3	0	0.308 6
	Living in center area	0.032 1	0.053 7	0	0	0

## 4 Conclusion

This paper demonstrates the underlying significance of developing a trip chaining pattern recognition model, and also explores the potential benefits of using the SVM model with its classification function to recognize trip chaining patterns. The SVM model shows better performance than the MNL model and the RBFNN model. In addition, the sensitivity analysis of the SVM model can evaluate the effects of explanatory variables on trip chaining behavior because the SVM model can produce more reliable results. Meanwhile, the contribution of different factors is summarized, which provides some thoughts on how to improve model performance.

Despite the insights observed and trials tested in this study, there is still some further research that needs to be considered. For instance, the enhancement to data availability and quality, and the extraction of effective variables can help complement and distinguish different trip chaining pattern characteristics. Some underlying truths should be excavated by considering several explanatory variables simultaneously if a more deliberate sensitivity analysis is conducted. A more sophisticated statistical method should be used for elaborate comparison. Future studies that consider these limitations will lead to more accurate forecasting for residents' trip chaining behavior, and will be helpful for transportation planning and management.

## References

- [1] Xianyu J C. An exploration of the interdependencies between trip chaining behavior and travel mode choice [J]. *Procedia-Social and Behavioural Sciences*, 2013, **96**: 1967 – 1975. DOI: 10.1016/j.sbspro.2013.08.222.
- [2] Li Z B, Wang W, Yang C, et al. Exploring the causal relationship between bicycle choice and trip chain pattern [J]. *Transport Policy*, 2013, **29**: 170 – 177. DOI: 10.1016/j.tranpol.2013.06.001.
- [3] Lee Y, Hickman M, Washington S. Household type and structure, time-use pattern, and trip-chaining behaviour [J]. *Transportation Research Part A: Policy and Practice*, 2007, **41**(10): 1004 – 1020. DOI: 10.1016/j.tra.2007.06.007.
- [4] Wang R. The stops made by commuters: Evidence from the 2009 US National Household Travel Survey [J]. *Journal of Transport Geography*, 2015, **47**: 109 – 118. DOI: 10.1016/j.jtrangeo.2014.11.005.
- [5] Ma J, Mitchell G, Heppenstall A. Daily travel behavior in Beijing, China: An analysis of workers' trip chains, and the role of socio-demographics and urban form [J]. *Habitat International*, 2014, **43**: 263 – 273. DOI: 10.1016/j.habitatint.2014.04.008.
- [6] Lee Y, Washington S, Frank L D. Examination of relationships between urban form, household activities, and time allocation in the Atlanta metropolitan region [J]. *Transportation Research Part A: Policy and Practice*, 2009, **43**(4): 360 – 373. DOI: 10.1016/j.tra.2008.11.013.
- [7] Frank L, Bradley M, Kavage S, et al. Urban form, trav-



- el time, and cost relationships with tour complexity and mode choice [J]. *Transportation*, 2008, **35**(1): 37 – 54. DOI: 10.1007/s11116-007-9136-6.
- [8] Maat K, Wee B V, Stead D. Land use and travel behavior: Expected effects from the perspective of utility theory and activity-based theories [J]. *Environment and Planning B: Planning and Design*, 2005, **32**(1): 33 – 46. DOI: 10.1068/b31106.
- [9] Shiftan Y. The use of activity-based modelling to analyse the effect of land-use policies on travel bahavior [J]. *The Annals of Regional Science*, 2008, **42**(1): 79 – 97. DOI: 10.1007/s00168-007-0139-1.
- [10] Li Z B, Liu P, Wang W, et al. Using support vector machine models for crash injury severity analysis [J]. *Accident Analysis and Prevention*, 2012, **45**: 478 – 486. DOI: 10.1016/j.aap.2011.08.016.
- [11] Xianyu J C. Travel mode choice analysis using support vector machine [C]//*11th International Conference of Chinese Transportation Professionals (ICCTP)*. Nanjing, China, 2011: 14 – 17.
- [12] Chang C C, Lin C J. LIBSVM: A library for support vector machines [EB/OL]. (2013-03-04) [2015-01-27] <https://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf>.
- [13] Powers D M W. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness & correlation [J]. *Journal of Machine Learning Technologies*, 2011, **2**(1): 37 – 63.
- [14] Cortez P, Embrechts M J. Using sensitivity analysis and visualization techniques to open black box data mining models [J]. *Information Sciences*, 2013, **225**: 1 – 17. DOI: 10.1016/j.ins.2012.10.039.

## 支持向量机在出行链模式识别和影响因素分析中的应用

杨 硕 邓 卫 程 龙

(东南大学交通学院, 南京 210096)

**摘要:** 为了提高交通需求预测精度, 综合考虑居民出行行为在时间维度上的分布, 采用支持向量机、径向基神经网络和多项 logit 三种方法, 基于居民活动目的, 建立了出行链模式识别模型, 并利用敏感性分析方法研究了解释因素对出行链模式选择的影响和对模型性能的贡献程度. 结果显示: 支持向量机模型在总体准确度和分类准确度上均优于其他 2 种方法, 体现了支持向量机在小样本下的识别性能优势; 证明了支持向量机能够较准确地反映多分类因素对于出行链模式选择行为的影响程度; 因素对于不同出行链模式识别精度的贡献度差异表明了细化出行链模式及探索各个模式特有影响因素的重要性. 支持向量机技术在交通需求预测建模及影响因素分析方面均具有实践意义.

**关键词:** 出行链模式; 支持向量机; 预测性能; 敏感性分析

**中图分类号:** U121