International Journal of Smart Computing and Artificial Intelligence
International Institute of Applied Informatics
2023, Vol. 7, No. 1, IJSCAI706

Evaluation of Zones in Himeji City for Foreign Tourists Using Support Vector Machines

Satoru Hakukawa *, Teijiro Isokawa *, Naotake Kamiura *

Abstract

In this paper, we present a method of evaluating zones in Himeji City, Japan, from the viewpoint of sight scene resources, to promote the tourism for foreigners visiting that city. Our method is based on the data classification conducted by support vector machines (SVMs for short). We prepare data presented to discrimination models constructed by SVM learning from tourist numbers totaled by country. In other words, the element value in the data equals the number of the tourists departing from some country and visiting some zone in Himeji City. Our model judges whether a zone of one square kilometer is worth a visit for the tourists departing from each of the following countries: France, United Kingdom (UK for short), Germany, Spain, Singapore, Australia, and United States of America (USA for short). It is established, from experimental results, that our method achieves substantially high averaged values on recall, precision, and F-measure when data to train our model are prepared from numbers of the tourists departing from six countries out of the above seven ones considered to be of importance in terms of tourism promotion.

Keywords: Foreign tourists, Himeji City, Support vector machines, Tourism evaluation

1 Introduction

The number of foreign tourists visiting Japan exceeded 30 million in fiscal 2018, and reached 31.88 million. It was the largest since the survey was first conducted by Japan Tourism Agency, in fiscal 2019 [1]. It has thus increased year after year, and induces the consumption in Japan. The consumption by foreign tourists in Japan was 4.8 trillion yen in fiscal 2019, and has increased for 8 consecutive years [2]. The Japanese Government has encouraged inbound tourism by foreigners as the promising field of the growth strategy. Himeji City, which is located in Hyogo Prefecture Japan, also established the plan in tourism and exchange promotion in 2007 [3], to activate inbound tourism. Himeji Convention Bureau has considered tourists departing from France, UK, Germany, Spain, Singapore, Australia, and USA to be open to the promotion, while Taiwan, South Korea, Thailand, and Hong Kong to be places reaching maturity in tourism to Himeji City [4].

To make fruitful promotion for inbound tourism to Himeji City, it is indispensable to ade-

* University of Hyogo, Hyogo, Japan

quately understand favorable zones with powerful tourism resources, and to give attractive information. It seems that the number of tourists departing from France, UK, Germany, Spain, Singapore, Australia, and USA, which are specially of importance for the promotion, can be one of the clues in evaluating the zones in terms of tourist attractions. The tourists visiting Himeji City from these seven countries, however, are not always majorities. Let us assume that the country to be investigated is determined as a target, and that thresholding is simply adopted on the number of tourists departing from the target to conduct the zone evaluation. The zone with extremely small number of the tourists departing from the target tends to be judged that it is worthless a visit. The judgement similar to the above would often be made for such zones, even if they have potentially strong attractive resources with which the tourists are unfamiliar.

In this paper, we present a method of evaluating zones from the viewpoint of tourism attractivity for tourists visiting Himeji City from a country for which the tourism promotion should be activated. We adopt data classification using SVMs [5], [6] in evaluating zones. Our method prepares the data for constructing discrimination models and executing the classification from the distribution on numbers of foreign tourists visiting Himeji City. These numbers were aggregated during one year and a half, and totaled by country. The numbers of tourists departing from 45 countries are available when our method prepares the data. We first choose a county to be explored as a target, and exclude the numbers of tourists departing from the target country. We then prepare data, each of which has numbers of tourists visiting one-square-kilometer zones in Himeji City from the remaining countries as element values. The number of elements is thus equal to that of all of the available countries minus one. An eight-digit number is assigned to each of the data. It specifies the position of the zone. In addition, our method assigns either *worthy* or *worthless* as a label to each of the training data to construct discrimination models. Assigning the labels to each of the training data depends on whether tourists departing from the country to be investigated actually visited the zone indicated by the eight-digit number assigned to each of them. We judge whether each of the test data with the eight-digit zone numbers belongs either to the *worthy* class or to the *worthless* class. If its class is judged to be *worthy*, our method considers the zone specified by its eight-digit number to be worth a visit for tourists departing from the target.

Let us assume that no tourists departing from the target country to be investigated actually visited some zone. Besides, let us assume that a large number of tourists departing from other counties actually visited that zone. If they have leisure taste and tourism tendency similar to the tourists departing from the target country, the corresponding zone is considered to be potentially worth a visit for the tourists departing from the target country. This intuition motivates us to apply our method to evaluate zones in Himeji City. It is established that our method achieves favorable metric values for the zone evaluation, provided that the number of countries (i.e., elements of data) is restricted to the number of counties from which Himeji Convention Bureau has considered tourists departing from the target country to be open to the promotion.

2 Spatial Mobile Data Used for Analysis

In this paper, we use spatial mobile data on the number of foreign tourists to analyze zones where they tend to visit in Himeji City. The data have been sold by DOCOMO InsightMarketing, INC [7]. In the dataset, the number of tourists is totaled by country. Entries in Table 1 are examples of the used data. The zone numbers in the leftmost column specify one-square-kilometer zones in Himeji City. In other words, such zones are considered to be statistically processing units. Note that the leftmost column includes information on season and year when the foreign tourists

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

Table 1: Examples of spatial mobile data

Season, year and zone number	Countries	Number of tourists
Summer, 2018, 52341584	USA	181
Summer, 2018, 52341584	China	685
⋮	⋮	⋮
Spring, 2019, 52343544	Korea	87

departed from each country for Japan visited the one-square-kilometer zone. Spring, summer, fall, and winter correspond to three months from March to May, those from June to August, those from September to November, and those from December to February, respectively. The second column is for the countries from which the tourists departed. The numbers in the rightmost column are for tourists visiting Japan from 45 countries. They are not cumulative total values. If some tourist has visited some zone many times in a single season, therefore, he/she is counted just once.

3 Tourism-Zone Evaluation Using Support Vector Machines

We construct discrimination models according to SVM learning, and evaluate one-square-kilometer zones in terms of tourism as part of developing the system to find new zones where the tourists departing from countries considered to be targets of tourism promotion should visit.

When determining year, season, and eight-digit number assigned to some zone, we obtain the numbers of the tourists coming from 45 countries. We consider the countries to be vector elements, and produce a set of 45-dimensional vectors. In other words, an element value corresponds to the tourist number for each of the countries. The lines in Table 2 show examples of such 45-dimensional vectors.

Let us next describe our evaluation for tourist zones in Himeji City. We first determine a target. The target means the country from which we wish to attract tourists departing. We next assign labels to the 45-dimensional vectors. For a vector, if the value of element corresponding

Table 2: Examples of 45-dimensional vectors

Season, year and zone number	1st element USA	2nd element Australia	…	45th element Korea
Summer, 2018 52341584	181	0	…	265
Summer, 2018 52341585	299	66	…	522
⋮	⋮	⋮	⋮	⋮
Spring, 2019 52343544	58	0	…	87

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

to the target is larger than zero, *worthy* is assigned to the vector as its label; otherwise, *worthless* is assigned. A vector has one-to-one correspondence with a one-square-kilometer zone specified by an eight-digit zone number as shown in Table 2. We consider the zone with the eight-digit number to be worth a visit for the tourists departing from the target, when the value of element corresponding to the target is not equal to zero. We finally remove the element associated with the target from each of the 45-dimensional vectors. This results in acquiring a dataset with 44-dimensional vectors as members. We analyze one-square-kilometer zones in terms of tourism, using this dataset. We present the members in the set for SVM learning. Figure 1 depicts the above preparation of the members for SVM learning, i.e., training data.

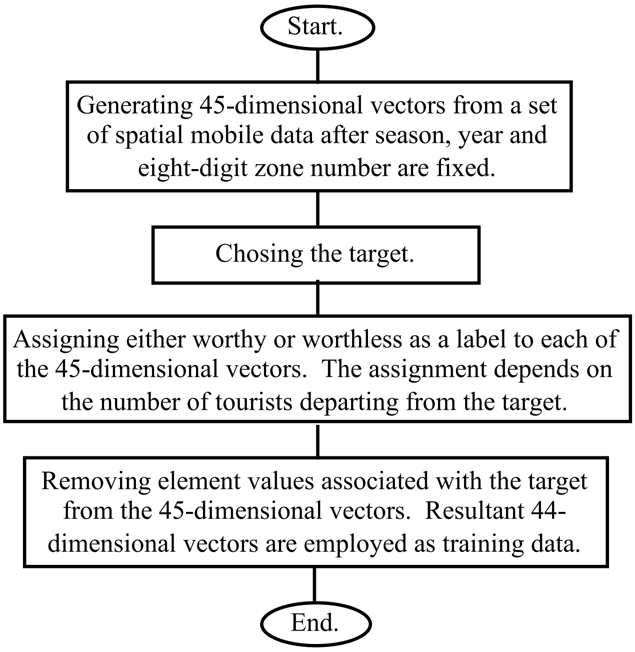

Figure 1: Preparation of training data

We apply RBFSVM learning algorithm [8]. It is available at scikit-learn [9]. When conducting SVM learning, we generally determine two parameters [10], C controlling the allowable extent of misclassification, and γ on which the complexity of the hyperplane in an input space depends. To determine values of C and γ, we apply a grid search based on five-fold cross-validation. SVM learning produces a discrimination model applicable to data classification.

It is considered that the input space consists of the area to which the test data with *worthy* labels should belong and the area to which the test data with *worthless* labels should belong. Note that the former area is strongly relevant to zones in Himeji City where the tourists departing from the target actually visited. Our method is based on the following intuition: for some zone in Himeji City, the tastes of the tourists departing from the target are potentially characterized by element values of the tourists visiting there from untargeted countries. In other words, when the class of some test data are judged to be *worthy*, we expect that such data have substantial visiting tendency and tastes associated with the tourists departing from the target. Results of data classification thus seem to be useful in newly evaluating zones potentially suitable for tourists departing from the target.

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

We prepare test data in a manner similar to preparing training data. In other words, no element associated with the target is excluded in the test data, and the test data have 44 element values. We also assign the eight-digit zone number to each of them. We check the classes of test data, using discrimination model constructed by SVM learning. When some test data are judged to be *worthy*, we conclude that zones specified by eight-digit numbers assigned to them are worth a visit for the tourists departing from the target. The above judgement is applied to every test data.

We finally visualize judgment results on the map of areas in Himeji City according to Folium [11]. It is expected that users can easily understand tendencies and characteristics of zones judged to be worth a visit.

4 Experimental Results

4.1 Classification-Based Zone Evaluation Using All Countries as Data Elements

When we first generate vectors from a set of spatial mobile data, the numbers of tourists departing from 45 countries are available. Himeji Convention Bureau has given priority to promoting tourism to Himeji City for people departing from France, UK, Germany, Spain, Singapore, Australia, and USA. We therefore choose one of these seven countries as a target, and construct discrimination models using training data prepared as shown in Figure 1. Note that each of training data has 44 element values. We then evaluate one-square-kilometer zones in terms of tourism attractivity for people departing from each of the seven countries chosen as the target.

In this subsection, spatial mobile data are numbers of tourists actually visiting Himeji City in summer (months from June through August 2018), fall (those from September through November 2018), winter (those from December 2018 through February 2019), and spring (those from March through May 2019). A set for training consists of 43 data associated with summer and fall, whereas that for test consists of 60 data associated with winter and spring. For evaluation, we estimate recall, precision, and F-measure for test data. Note that recall is the ratio of the number of zones accurately judged to be worth a visit for the tourists departing from the target compared to that of zones where the tourists departing from the target actually visited. Assume *NZV* to be the number of zones judged to be worth a visit for the tourists departing from the target, and *NZVR* to be the number of zones where the tourists departing from the target actually visited among the *NZV* zones. We then have *NZVR*/*NZV* as precision. Besides, we have F-measure by calculating $2 \times rec \times prec \,/\, (rec + prec)$, where *rec* and *prec* denote the recall and precision values, respectively.

We show experimental results in Table 3, Figures 2 and 3. These figures are for examples of map-based zone evaluation when we choose France and Australia as targets, respectively. A red circle hereinafter expresses a one-square-kilometer zone judged to be worth a visit for one of the test data presented to the constructed discrimination model. We can consider that each of the test data has a season and a year (i.e., winter 2018 or spring 2019) as a label, in addition to an eight-digit zone number. Two test data with the same zone number are thus available. Note that a circle has two pins at most in Figures 2 and 3. If a red circle has two pins, it is judged to be worth a visit for the tourists departing from the target for test data both with winter 2018 label and with spring 2019 label. When the number of pins is one, this is a case where our method judges the circle to be *worthy* class either for test data with winter 2018 label or test data with spring 2019 label. The colors of pins have the following meanings. If the color is blue (or pink), it means that our method judges the circle where tourists departing from the target actually vis-

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

ited in winter 2018 (or spring 2019) to be worth a visit by presenting test data with winter 2018 (or spring 2019) label. The meaning of a red pin is as follows: there are no tourists that actually visited the circle with this pin from the target in winter 2018 (or spring 2019), though our method judges it to be worth a visiti for test data with winter 2018 (or spring 2019) label. Some zones judged to be worth a visit appear around JR Himeji station in Figure 2. In addition, the real area with circles shown in Figure 3 is much wider than that shown in Figure 2.

Table 3: Numerical results when 44-dimensional data are employed

Target countries	Evaluation metrics		
	Recall	Precision	F-measure
France	1.00	0.69	0.82
UK	0.60	1.00	0.75
Germany	0.30	1.00	0.46
Singapore	0.60	1.00	0.75
Spain	0.71	0.83	0.77
Australia	1.00	0.23	0.37
USA	1.00	0.42	0.59
Average	0.80	0.45	0.58

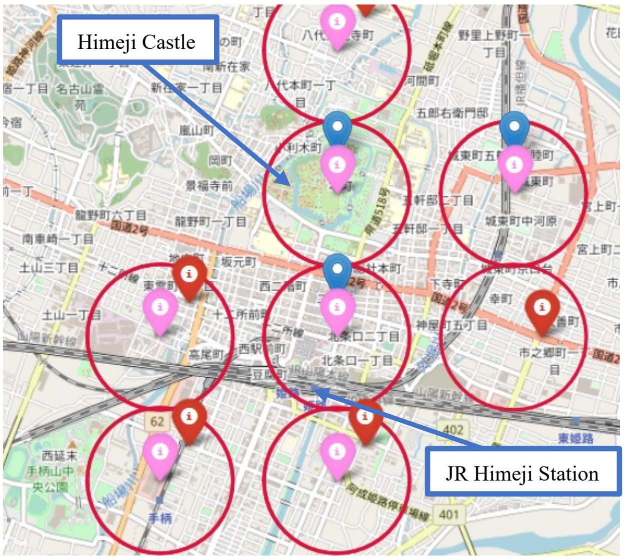

Figure 2: Map-based zone evaluation for tourists departing from France

We have 0.45 for the precision value on average as shown in Table 3, whereas averaged recall is equal to 0.80. Averaged F-measure equal to 0.58 is also somewhat disappointing. On the other hand, we achieve favorable recall and F-measure values for tourists departing from France. Five circles with red pins appear in Figure 2. Recall that, in the case where red pins are dropped, no tourists departing from the target visited the circles with such pins either in winter 2018 or in spring 2019. It is therefore probable to construe that our method newly evaluates the five circles to be worth a visit for the tourists departing from France. We next discuss the cases where targets are UK, Germany, and Singapore. Recall values are equal to 0.60, 0.30, and 0.60 for them, respectively, whereas we have the value of 1.00 as the precision values. Low recall and high precision tend to be simultaneously held when we judge circles to be worthless a visit, though the

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

target tourists actually visited them. For tourists departing from either Australia or USA, F-measure values are low. Especially, on Australia, we judge all circles specified by zone numbers attached to all test data to be worth a visit, as shown in Figure 3.

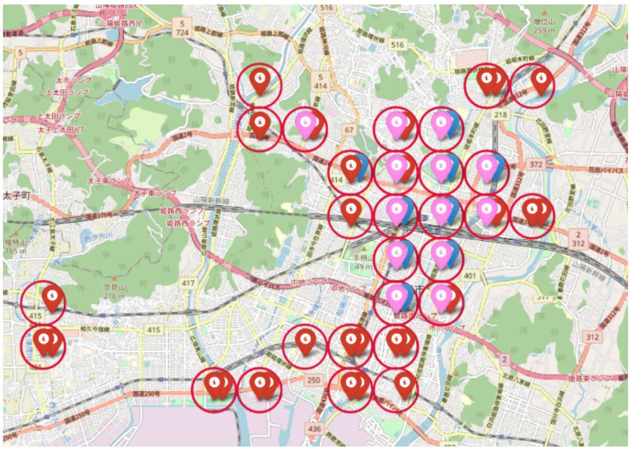

Figure 3: Map-based zone evaluation for tourists departing from Australia

As discussed in Section 3, a member of the training dataset and test dataset has 44 element values corresponding to numbers of tourists departing from 44 countries. The difference in the tourist number by country would cause somewhat disappointing numerical results in Table 3. Entries in Table 4 are tourist numbers for the seven countries to be chosen as targets, whereas those in Table 5 are numbers of tourists departing from bottom ranked five countries. They were aggregated from summer 2018 through spring 2019, and used as element values in training data and test data.

Table 4: Numbers of tourists departing from seven countries chosen as targets for a year

Countries	Number of tourists
USA	26715
France	20316
Australia	16096
Spain	10545
UK	8501
Germany	6922
Singapore	5452

Table 5: Numbers of tourists coming from bottom five countries not chosen as targets for a year

Countries	Number of tourists
Croatia	68
Latviya	71
Estonia	88
Turkey	96
Luxembourg	139

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

The quite large difference exists between each of the entries in Table 4 and that in Table 5. Himeji Convention Bureau tends to exclude countries from targets for the promotion, if it is unexpected that a very large number of tourists departing from them visit Himeji City. It is undeniable that training data would affect the capability of discrimination models, if they have some elements having extremely small element values. It is probable that such elements correspond to countries that Himeji Convention Bureau often discusses not to be of importance in promoting tourism. We therefore discuss the case of reducing elements (i.e., countries) in training data and test data.

4.2 Classification-Based Zone Evaluation Using Six Countries as Data Elements

We prepare training data and test data, provided that elements in them are restricted to France, UK, Germany, Spain, Singapore, Australia, and USA that Himeji Convention Bureau considers to be of importance. Except for the restriction of the number of elements, the manner in preparing the data is similar to that mentioned in Section 3. We choose the target. Since the target is one of the above seven countries, the data have six elements corresponding to the remaining countries. The data are also labeled. In other words, season, year, and eight-digit zone number are also assigned to each of them.

We first discuss the case where a training dataset is prepared from spatial mobile data aggregated in summer 2018 and in fall 2018, while a test dataset is from data aggregated in winter 2018 and in spring 2019. In other words, members in the former set have either summer 2018 or fall 2018 as labels, and those in the latter set have either winter 2018 or spring 2019. We tabulate the evaluation results for one-square-kilometer zones in Table 6. In addition, Figure 4 is for the map-based zone evaluation applied to tourists departing from Australia.

Table 6: Numerical results when six-dimensional data are employed

Target countries	Evaluation metrics		
	Recall	Precision	F-measure
France	1.00	0.85	0.92
UK	0.60	0.86	0.71
Germany	0.40	1.00	0.57
Singapore	0.90	0.90	0.90
Spain	0.86	0.75	0.80
Australia	0.93	0.65	0.77
USA	1.00	0.42	0.59
Average	0.85	0.61	0.71

As shown in Table 6, our method achieves 0.71 as averaged F-measure value. It is high, compared to averaged F-measure value achieved in the case of employing the dataset with 44-dimensional vectors as members. Note that the results in this case are tabulated in Table 3. Our method also achieves high averaged values for the other metrics. It seems that setting the number of elements to six results in the improvement of averaged metric values. We next discuss the results for every country chosen as the target. When the target is France, we achieve precision and F-measure values higher than those shown in Table 3. In the case of choosing UK as the target, our method employing six-dimensional data makes the precision value worse, whereas preserving the recall value. These precision and F-measure values cause the slight

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

degradation of F-measure value. For tourists departing from Spain and Singapore, since recall values increase and precision values decrease, our method improves F-measure values. Let us next discuss for tourists departing from Australia. When our method employs six-dimensional data, the recall value slightly deteriorates. It, however, allows us to improve precision and F-measure values. Besides, our method can decrease the number of red circles judged to be worth a visit as shown in Figure 4, compared to that in Figure 3. This result implies that we can avoid unhelpful cases of evaluating almost all of the zones in Himeji City to be *worthy*. For tourists departing from USA, all metric values achieved using 44-dimensional data are equal to those achieved using six-dimensional data. For tourists departing from Germany, while recall and F-measure values are improved, they are somewhat low. Our method must rise the three metric values on Germany more.

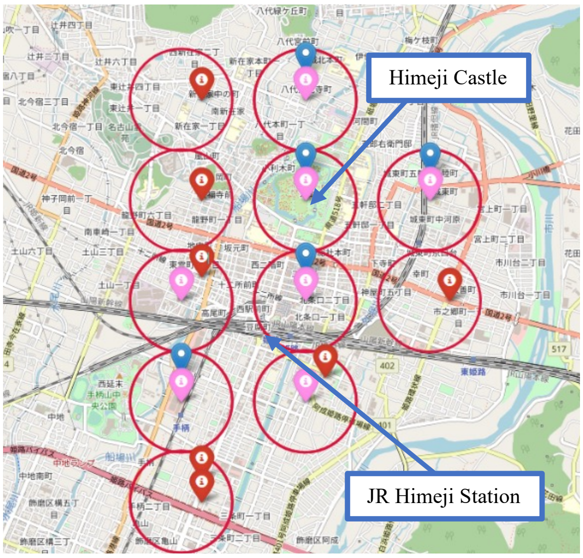

Figure 4: Map-based zone evaluation for tourists departing from Australia, subject to employing six-dimensional training and test data

We next discuss the case of increasing the number of training data. We prepare six-dimensional training data from a set of spatial mobile data aggregated for a year (from summer 2018 through spring 2019). The number of training data is then equal to 103. Besides, we use 73 test data prepared from spatial mobile data aggregated in summer 2019 and in fall 2019. The results are tabulated in Table 7. The averaged F-measure value in Table 7 is slightly preferable to that in Table 6. The averaged recall and precision values are also higher than those in Table 6. Let us next discuss the results for each of the six targets. For tourists departing from France, all metric values in Table 7 are unfavorable compared with those in Table 6. On the other hand, when UK and Australia are chosen as targets, all metric values in Table 7 are higher than those in Table 6. In the cases where Spain, Germany, and Singapore are chosen, our method achieves higher F-measure values as shown in Table 7 than in Table 6, whereas either recall value or precision value in the former is lower than that in the latter. When USA is chosen, the unfavorable evaluation judging almost all of the zones in Himeji City to be *worthy* occurs, whereas all metric values in Table 7 are as high as those in Table 6. Note that this unfavorable evaluation is similar to the map-based zone evaluation in Figure 3.

We next discuss seasonality by applying discrimination models constructed with training data

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

having summer and fall 2018 as their labels to test data having summer and fall 2019. In this case, the number of training data (or test data) is 43 (or 73). The results are tabulated in Table 8. All averaged metric values in Table 8 are slightly lower than those in Table 6. We focus on the results for each of the six targets. When the target is France, all metric values become low compared to those in Table 6. On the other hand, for tourists departing from UK, all metric values become large compared to those in Table 6. In the cases where Spain, Australia, and Singapore are chosen as targets, our method achieves higher values either for recall value or for precision than those in Table 6. The F-measure value for tourists departing from Spain is higher than that in Table 6, and the F-measure value is equal to that in Table 6 for tourists departing from Australia. It however is lower than the F-measure value in Table 6 for tourists departing from Singapore. When Germany and USA are chosen, disappointing metric values appear in Table 8 as well as in Table 6.

Table 7: Numerical results when six-dimensional training data for a year are employed

Target countries	Evaluation metrics		
	Recall	Precision	F-measure
France	0.82	0.64	0.72
UK	0.77	1.00	0.87
Germany	0.85	0.92	0.88
Singapore	1.00	0.86	0.92
Spain	0.73	1.00	0.83
Australia	1.00	0.80	0.89
USA	1.00	0.44	0.61
Average	0.90	0.63	0.74

Table 8: Numerical results of experiments in consideration of seasonality

Target countries	Evaluation metrics		
	Recall	Precision	F-measure
France	0.91	0.67	0.77
UK	0.69	1.00	0.82
Germany	0.31	1.00	0.47
Singapore	1.00	0.73	0.84
Spain	0.73	1.00	0.84
Australia	1.00	0.63	0.77
USA	1.00	0.44	0.61
Average	0.83	0.59	0.69

Let us next discuss the case of increasing the number of training data and employing the test dataset with members to which winter 2018 and spring 2019 are assigned as season and year labels. Note that the test dataset is then identical to that used to obtain numerical results in Table 6. The training data are prepared from spatial mobile data aggregated in summer and fall 2018 and in summer and fall 2019. The number of them is 116, whereas that of test data is 60. The numerical results are tabulated in Table 9. The averaged precision and F-measure values are higher than those in Table 6, whereas remaining averaged recall value is lower than that in Table 6. For tourists departing from UK, Germany, Australia, and USA, it can be comprehensively considered that our method can achieve favorable metric values compared to those in Table 6. On the other hand, metric values for France, Singapore, and Spain are somewhat inferior to those

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

in Table 6. No extremely low metric values, however, appear in Table 9, though such values are shown on lines for Germany and USA in Table 6. The map-based zone evaluation when we choose USA as the target is depicted in Figure 5. Note that the unfavorable evaluation judging almost all of the zones in Himeji City to be *worthy* occurs in Figure 3. Such unfavorable evaluation also occurs in the experiment bringing about numerical metric values in Table 7 when the target is USA. Since the number of zones judged to be worth a visit is much smaller than that on Figure 3, it is shown that we can obtain the meaningful result on USA.

Table 9: Numerical results when six-dimensional training data for four seasons in two years are employed

Target countries	Evaluation metrics		
	Recall	Precision	F-measure
France	1.00	0.79	0.89
UK	0.70	0.88	0.78
Germany	0.70	0.78	0.74
Singapore	0.70	1.00	0.82
Spain	0.86	0.67	0.75
Australia	0.86	0.92	0.89
USA	0.64	1.00	0.78
Average	0.76	0.87	0.81

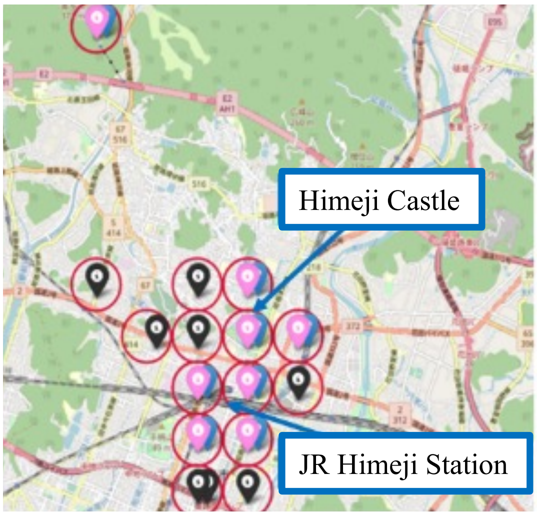

Figure 5: Map-based zone evaluation for tourists departing from USA

5 Discussion

We discuss tourists departing from USA subject to employing six-dimensional training data and test data, as an example of the case where metric values are unfavorable. Precision values in Tables 6-8 are low for the tourists departing from USA, compared to those for tourists departing from any other country. Low precision values tend to be caused when our method often misjudges one-square-kilometer zones where no tourists departing from the target visited to be *worthy*. For experiments bringing about the results tabulated in Table 6, entries in Table 10 are equal to the numbers of zones corresponding to training data (i.e., zones specified by eight-digit

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

numbers assigned to them). They are summed up by country. Note that the number of the zones associated with USA is the largest, and such zones are widely distributed in Himeji City. This implies that tourists departing from USA exclusively visited a considerable number of zones. The discrimination model constructed when the target is USA then tends to frequently judge the class of test data with the values of 0 for almost all of the elements to be worth a visit. Increasing the number of training data makes it possible to improve the precision value for tourists departing from USA as shown in Table 9. Except for increasing the number of training data, the promising approach to avoid errors, which affect forming the approximate input space corresponding to *worthy* class during SVM learning, seems to apply data cleansing to training data with the values of 0 for almost all of the elements.

Table 10: Number of zones specified by eight-digit numbers assigned to training data employed for experiments bringing about results in Table 6

Target countries	Number of zones
France	7
UK	4
Germany	6
Singapore	6
Spain	4
Australia	10
USA	13

An example of map-based zone evaluation is depicted in Figure 6 for tourists departing from UK, provided that six-dimensional training data and test data bringing about numerical results in Table 6 are employed. Our method gives black circles with black pins to zones judged to be *worthless* though the tourists departing from UK actually visited either in winter 2018 or in spring 2019. It can be then considered that such black circles express inappropriate evaluation. The black circles are located in the direction of south-west from JR Himeji station. Japanese tourists usually go along the sightseeing route connecting between JR Himeji station and Himeji castle, which is located in the direction of north from JR Himeji station. If tastes of Japanese tourists resemble those of foreign tourists on sightseeing, the number of foreign tourists visiting the south-west zones from JR Himeji station seems to be smaller than that of foreign tourists visiting the north zones. We next explore the relevance potentially lying between evaluation results of the zones and numbers of the tourists actually visiting there. We tabulate zones inappropriately (or appropriately) evaluated to be worthless (or worth) a visit and numbers of tourists actually visited there in Table 11 (or 12) for some of the targets. Note that the zones in Tables 11 and 12 are evaluated by employing datasets to conduct experiments bringing about results in Table 6.

It is revealed, from Tables 11 and 12, that our method often misjudges zones where tourists departing from targets actually visited to be worthless a visit as the numbers of such tourists become smaller. Since our discrimination model is basically designed so that it can be available for two-class classification problems, we must assign *worthy* labels to training data, so long as at least one tourist departing from the target visits corresponding one-square-kilometer zones. This labeling is made regardless of the absolute values of the numbers of tourists actually visiting the corresponding zones. The above misjudgments for the zones where extremely small number of tourists departing from targets actually visited seem to be caused by preparing the data in this manner. It is considered that developing the multi-class classification, which depends on data

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

labeling making much account of difference in numbers of tourists actually visited, is one of the promising approaches to overcome this problem.

Let us qualitatively explain the limitations of our method. Recall that our method can evaluate just one-square-kilometer zones in Himeji City from the viewpoint of tourism. It is impossible to apply our method to a tourism resource of which the area is less than one square kilometer. For the zone evaluated by our method to be worth a visit, except for the inference made by users, there are no ways to determine what resources, which are located in the zone, the tourists departing from a target consider to be attractive. This would cause a problem in choosing tourism resources to be promoted.

As described in Section 2, spatial mobile data are numbers of tourists departing from 45 countries, and our method can prepare training data and test data from them. In other words, if the country to be explored as a target is none of these 45 countries, our method is absolutely unapplicable. It is next impossible to precisely know that the main object of visiting Himeji City is either a sightseeing tour or a business trip, from the spatial mobile data. In almost all the cases of visiting Himeji City, it seems to be sightseeing. It is very difficult to simply apply our method to metropolitan areas where the number of business trips is almost equal to that of sightseeing tours, in terms of tourism promotion. However, after the spatial mobile data are perfectly divided into business-trip dataset and sightseeing-tour dataset in some manner, if training data are prepared from members belonging to the latter dataset, it is expected that our method can conduct the highly precise evaluation for one-square-kilometer zones.

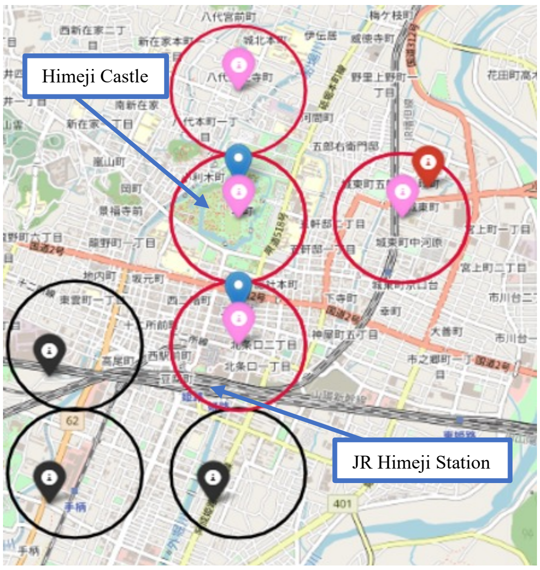

Figure 6: Map-based zone evaluation for tourists departing from UK

6 Conclusions

In this paper, we proposed the method of evaluating zones in Himeji City for tourists departing from countries chosen as targets in terms of sightseeing. Our method is based on discrimination models constructed by SVM learning employing data prepared from spatial mobile data with 45 elements. The countries from which foreign tourists visited Himeji City correspond to these elements. The training data have seasons, years in which the spatial mobile data are aggregated, and eight-digit zone numbers as their labels. We investigated two types of the evaluation speci-

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

Table 11: Examples of zones inappropriately evaluated to be worthless a visit and numbers of tourists actually visited there

Target countries	Zone name, Season	Number of tourists
UK	Shoshazan, Spring	119
	Chiyodacho, Spring	88
Germany	Honcho, Winter	110
	Chiyodacho, Winter	61
Spain	Shoshazan, Spring	107
Singapore	Hojyo, Winter	95
Australia	Shoshazan, Winter	76

Table 12: Examples of zones appropriately evaluated to be worth a visit and numbers of tourists actually visited there

Target countries	Zone name, Season	Number of tourists
UK	Honcho, Spring	1748
Germany	Honcho, Spring	1800
Spain	Jyotocho, Winter	499
Singapore	Higashiekimaecho, Spring	946
Australia	Honcho, Winter	1058

ied by the element number on prepared data. The first evaluation employs the dataset with 44-dimensional members, whereas the second evaluation was made by presenting the six-dimensional data to discrimination models. The six elements in the data for the second evaluation come from countries that Himeji Convention Bureau considers to be specially promising for the tourism promotion. In the best case of the second evaluation, our method achieves 0.81 as a F-measure value on average. It was therefore revealed that restricting the number of elements and increasing the number of training data result in raising the accuracy of zone evaluation.

In future, our method will be modified so that higher metric values can be achieved, compared to those achieved at present, by introducing data cleansing and the multi-class classification.

Acknowledgement

This research was supported in part by JSPS KAKENHI Grant-in-Aid for Scientific Research (C) (Grant number JP21K12458).

References

[1] Numbers of tourists visiting Japan and Japanese Nationals Departing from Japan, https://www.mlit.go.jp/kankocho/siryou/toukei/in_out.html (in Japanese), Dec. 27, 2021 (Final reference time).

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.

[2] The value of travel consumption of inbound tourists in fiscal 2019, https://www.mlit.go.jp/kankocho/siryou/toukei/content/001335741.pdf (in Japanese), Dec. 27, 2021 (Final reference time).

[3] https://www.mlit.go.jp/common/001270416.pdf (in Japanese), pp.2-31-2.36, Dec. 27, 2021 (Final reference time).

[4] Business Policy of Public Interest Incorporated Association, Himeji Convention Bureau for Tourism in fiscal 2019, https://himeji-kanko.jp/pdf/bureau/r01-keikaku.pdf (in Japanese), Dec. 27, 2021 (Final reference time).

[5] V. N. Vapnik, Statistical Learning Theory, Wiley, New York, p.281, 1998.

[6] C. Cortes and V. Vapnik,"Support Vector networks," Machine Learning, Vol. 20, pp. 273-297, 1995.

[7] On Practical Application of Spatial Statistics for Mobile Terminals, https://www.nttdocomo.co.jp/binary/pdf/info/notice/page/130906_00.pdf (in Japanese) Dec. 27, 2021 (Final reference time).

[8] T. Onoda, Science of the Intellect - Support Vector Machine (in Japanese), Ohmsha, Ltd., p. 130, 2007.

[9] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research 12, 2011, pp.2825-2830.

[10] C. C. Chang and C-J. Lin, "LIBSVM: A Library for Support Vector Machines," ACM Transactions on Intelligent Systems and Technology, https://doi.org/10.1145/1961189.1961199, Vol. 2, No. 3, 2011, pp.3-5.

[11] R. Story, "Folium," https://python-visualization.github.io/folium, 2013, (Dec. 27, 2021 : Final reference time).

Copyright © by IIAI. Unauthorized reproduction of this article is prohibited.