Fishing Spot Prediction by Sea Temperature Pattern Learning

Masaaki Iiyama
Kyoto University
Kyoto, Japan

Kei Zhao
Kyoto University
Kyoto, Japan

Atsushi Hashimoto
Kyoto University
Kyoto, Japan

Hidekazu Kasahara
Kyoto University
Kyoto, Japan

Michihiko Minoh
Kyoto University
Kyoto, Japan

Abstract—Determination of appropriate fishing spots is one of the most important activities in the fishing industry. Inspired by the approach followed by fishermen to determine fishing spots, this paper presents a new machine-learning method for uncovering oceanographic patterns related to good fishing spots. Our method uses a sea temperature map as the input, extracts sea temperature patterns from arbitrary points on the map, and evaluates whether the patterns correspond to good fishing spots by using two machine learning techniques; one-class support vector machine (SVM) and spectral clustering. We evaluated the efficiency of our method using fishery data on neon flying squid.

I. INTRODUCTION

Determining good fishing spots improves fishing efficiency and is one of the most important activities in the fishing industry. Recent technologies of remote sensing by satellites and data assimilation are able to describe and predict the marine environments of the ocean. Since the marine environment is an important clue to estimate fishing spots in fisheries science, many attempts have been made to discover the mechanism of the ecology of fish by modeling the Habitat Suitability Index based on the marine environment [1].

We focus on fishermen’s behaviors when locating fishing spots, whereas such conventional research in fisheries science focuses on fish ecology. Fishermen use ocean condition maps (e.g., sea temperature maps) for deciding where to go fishing. They find specific structures of the ocean, such as warm water eddies, which would be highly related to good fishing spots. From the viewpoint of informatics, this procedure can be modeled as a pattern recognition problem. Inspired by the fishermen’s behaviors, we analyzed the relationship between oceanographic patterns and fishing spots and then employed a machine-learning approach to uncover such patterns related to good spots.

An overview of our approach is illustrated in Fig.1. We take a sea temperature map as the input and then extract the sea temperature patterns from arbitrary points on the map. For each pattern, we evaluate whether the patterns will be potentially good fishing spots and suggest these spots to fishermen. The evaluation is automatically conducted by a pattern classifier that is constructed by our machine-learning approach.

The machine learning approach requires the training data to construct the classifier. We assumed two different scenarios for this approach.

II. ONE-CLASS SVM FOR LOCATING POTENTIAL FISHING SPOTS

One possible solution for determining potential fishing spots is to construct a two-class classifier that takes an unseen sea temperature pattern as the input and outputs a binary (0 or 1) value, which indicates whether the inputted pattern signifies a suitable fishing spot. The locations of already known spots, that is, locations that fishermen have recognized as good spots, are used as *positive* samples. The problem here is the lack of explicit *negative* samples, that is, locations not suitable for fishing. Fishermen go only to the locations they believe to be good spots. Hence, we cannot treat the unvisited locations as negative samples. In order to solve this problem, we employ a

Fig. 1. Using a machine-learning approach to determine potential fishing spots

The first scenario is a case that past sea temperature patterns being paired with corresponding logbook entries (dates, locations of fishing spots, and fish catches) are available. When a large number of logbooks are available, we can employ a supervised learning approach.

The second scenario is a case that logbooks are not available but past fishing spots are available. For example, the locations of fishing boats can be acquired from night satellite images or Automatic Identification System (AIS) or Vessel Monitoring System (VMS). A research project was started to collect AIS and VMS data from fisheries[1]. In such cases, we cannot employ the supervised learning approach but can employ an unsupervised approach.

We propose the following methods for both scenarios.

1. Global Fishing Watch (http://globalfishingwatch.org/ )
one-class support vector machine (SVM) [2], which requires only positive samples for training the classifier.

From the logbooks, we extract a sea temperature pattern as a cropped rectangular centered on the location of a boat at a recorded date on the sea temperature map. We collect the data from past logbooks in order to train the one-class SVM. Since the sea temperature pattern is a 2D array of high dimensional data, we applied kernel-PCA for dimension reduction and an acquired vector \( \mathbf{x} \) to represent the sea temperature pattern. We call \( \mathbf{x} \) the sea temperature feature. Sea temperature patterns extracted from the training data are used for calculating the principal components. Feature \( \mathbf{x} \) of a given sea temperature pattern acts as the coefficient of the principal components that approximate the given sea temperature pattern.

The original SVM is a classifier that finds the decision boundary to discriminate between two classes. However, the one-class SVM is used mainly for detecting anomalies, and it tries to find a decision boundary that locates \((1 - \nu)\) of the training data at one side of the decision boundary, where \(\nu\) is a parameter that represents the fraction of outliers in the training data. We used a fixed value for \(\nu (\nu = 0.0595)\) for further analysis. The classifier trained by the one-class SVM returns a likelihood that the sea temperature feature \( \mathbf{x} \) belongs to a positive class. We use this likelihood as a suitability for fishing spot.

### III. Clustering for estimating fishery catch

When data on past fish catches with regard to sea temperature patterns are available, another approach is applicable. One naive method is to construct a regression model that takes the sea temperature feature as the input and predicts the fish catches at the locations of the patterns. However, this does not work because similar patterns (similar features) do not always give similar fish catches. Actually, fish catch is affected not only by sea temperature patterns but also by many other factors, such as total fish resource, the skills of the fishermen. Thus, predicting fish catches from only sea temperature maps is difficult.

Instead of predicting fish catches using regression, we propose a classification-based approach. Here, we make an assumption about sea temperature patterns. The patterns at fishing spots are categorized into several typical patterns and those patterns at unsuitable locations do not belong to any categories. Some categories may supply a larger amount of fish catches than others. Our method discovers the categories and ranks them according to the expectancies of fish catches after determining the category of the input pattern.

Suppose that pairs of sea temperature feature \( \mathbf{x} \) and corresponding fish catch \( \mathbf{y} \) are available as training data. The features are acquired from the sea temperature patterns in the same manner as described in Section II. Then, we apply a clustering method for categorizing the features.

We expect that several typical sea temperature patterns suitable for fishing do exist and form individual clusters. However, actual training data contain outliers that are not suitable for fishing. For example, the logbooks of fishing boats may contain mistakes and records of some boats having visited unsuitable fishing spots. Such data are treated as outliers.

In order to cope with the outliers, we apply outlier-detectable spectral clustering [3] to the set of \( \mathbf{x} \) and categorize each \( \mathbf{x} \) into a cluster. Spectral clustering uses a similarity matrix \( S \) that represents the similarity between two sea temperature features. Our method employs the cosine similarity as a similarity metric. By using the spectral clustering, outliers form a single outlier cluster, whereas the inliers form several clusters. The data categorized into the outlier cluster represent unsuitable fishing spots.

For each cluster \( c_i \), we select the samples that belong to the cluster \( c_i \), and calculate the median of \( \mathbf{y} \) of the selected samples. We rank the clusters according to the median of \( \mathbf{y} \). When a sample is detected as an outlier, it would be a pattern obtained from poor fishing spots.

Once all \( \mathbf{x} \)'s in the training data are assigned to one of the clusters, we employ an SVM to construct a multi-class classifier that takes \( \mathbf{x} \) as the input and outputs the cluster ID to which \( \mathbf{x} \) is assigned (or recognizes \( \mathbf{x} \) as an outlier).

### IV. Experimental results

We evaluated our methods with data on fish catches. For training data, we used the logbooks of boats fishing for neon flying squid. Each sample contained the locations of fishing spots, dates, and amounts of fish catches. We extracted \(4^\circ \times 4^\circ\) pattern from a sea temperature map of 100m depth obtained from a data assimilation product [4].

#### A. Results with One-Class SVM

In order to evaluate the method described in Section II, we synthesized negative samples by extracting areas where most fishing boats do not stop. We used data from the years 1999 to 2006 for training and data from 2007 for evaluation. Figure 2 shows a ROC curve for the two-class fishing spot classification. We tested patterns of various sizes. Figure 2 indicates that there are no significant differences among the
pattern sizes, $3^\circ \times 3^\circ, 4^\circ \times 4^\circ$ and $5^\circ \times 5^\circ$, but also indicates that our pattern-based approach outperforms traditional point based approaches. This result implies that fishing spots are formed not only by the sea temperature at this point but also by the surrounding sea temperature patterns.

**B. Results with Clustering**

We used 5,058 samples from 1999 to 2009 for training and 1,181 samples from 2010 to 2012 for evaluation. Each sample contained the locations and dates of fishing, fish catches (CPUE: Catch Per Unit Effort), and corresponding sea temperature patterns of 100m depth.

Applying the method described in Section III, we acquired five clusters. We sorted these clusters according to the median of CPUE in a descending order and labeled them $c_1, \ldots, c_5$. The outlier cluster described in Section III is labeled $c_4$. The second column in Table I shows the median of the CPUE of the training data. Compared with $c_1$ and $c_5$, $c_1$ and All imply that our method successfully finds good (and bad) sea temperature patterns for fishing.

We evaluated the efficiency using the evaluation samples. The results are shown in the third and fourth columns of Table I. The median of the CPUE between 2010 and 2012 (1,181 samples) was 2.1. Our method classified 127 of the 1181 samples into the top-rank cluster $c_1$, and found the median to be 2.9. In contrast, 112 of the 1181 samples are classified into the worst cluster $c_5$ with a median of 1.5. These results indicate that our method ranks fishing spots successfully.

<table>
<thead>
<tr>
<th>Cluster $c_i$</th>
<th>'99-'09 CPUE</th>
<th>'10-'12 CPUE</th>
<th>'10-'12 # of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_1$</td>
<td>5.2</td>
<td>2.9</td>
<td>127</td>
</tr>
<tr>
<td>$c_2$</td>
<td>3.41</td>
<td>2.1</td>
<td>335</td>
</tr>
<tr>
<td>$c_3$</td>
<td>2.92</td>
<td>2.3</td>
<td>379</td>
</tr>
<tr>
<td>$c_4$</td>
<td>2.9</td>
<td>1.79</td>
<td>228</td>
</tr>
<tr>
<td>$c_5$</td>
<td>2.85</td>
<td>1.5</td>
<td>112</td>
</tr>
<tr>
<td>All</td>
<td>3.38</td>
<td>2.1</td>
<td>1181</td>
</tr>
</tbody>
</table>

Table I: Clustering results and medians of training and testing data

Figure 3 shows the predicted fishing spots and sea temperature maps of 100m depth as of 1st July 2012. Figure 3 (a) shows the suitability map obtained by one-class SVM method described in Section II. The yellow color indicates high suitability. Figure 3 (b) shows the fishing spot map by using spectral clustering based method described in Section III. Every point on the map is categorized into one of five clusters, $c_1, \ldots, c_5$. Points categorized into the outlier cluster $c_4$ are indicated by violet (worst suitability). Both Fig.3 (a) and (b) show similar potential fishing spots. Our spectral clustering method successfully finds unsuitable fishing spots.

Fig. 4 shows the results of five successive days, from 1 to 5 July 2011. The fishing spot moves slightly each day as the temperature pattern changes.

**V. Conclusion**

In this paper, we proposed a pattern analysis approach for estimating good fishing spots. We employed a one-class SVM
in order to detect fishing spots and clustering in order to estimate fishery catches. Although we tested our methods with only data of catches of neon flying squid, our methods can be applied to other fish. Applying the methods to other fish remains a topic for as future studies.

ACKNOWLEDGMENTS

We wish to thank Fisheries Research Institute, Aomori Prefectural Industrial Technology Research Center for providing us their expert advice and fishery data, and we also thank Meteorological Research Institute and JAMSTEC for providing oceanographic data. This work was supported by JST CREST Grant Number JPMJCR1681, Japan.

REFERENCES


