
Solving the oil spill problem using a combination of CBR and a summarization of SOM ensembles.

Aitor Mata¹, Emilio Corchado² and Bruno Baruque²

¹ University of Salamanca, Spain

² University of Burgos, Spain

aitor@usal.es, escorchado@ubu.es and bbaruque@ubu.es

Abstract . In this paper, a forecasting system is presented. It predicts the presence of oil slicks in a certain area of the open sea after an oil spill using Case-Based Reasoning methodology. CBR systems are designed to generate solutions to a certain problem by analysing historical data where previous solutions are stored. The system explained includes a novel network for data classification and retrieval. Such network works as a summarization algorithm for the results of an ensemble of Self-Organizing Maps. This algorithm, called Weighted Voting Superposition (WeVoS), is aimed to achieve the lowest topographic error in the map. The WeVoS-CBR system has been able to precisely predict the presence of oil slicks in the open sea areas of the north west of the Galician coast.

Keywords: Case-Based Reasoning; oil spill; Self Organizing Memory; summarization; Radial Basis Function.

1 Introduction

After an oil spill, decisions must be taken fast and accurately in order to avoid natural disasters. Forecasting the probability of finding oil slicks after an oil spill in a certain area will offer a great support to take those critical decisions. In this paper a forecasting system is presented, which aim is to proportionate the probability of finding oil slicks in a certain area using the Case-Based Reasoning methodology and a new summarization algorithm for SOM ensembles.

The system explained here has been developed using historical data and the information obtained after the Prestige accident in November 2002. To obtain data about the oil slicks, like their positions and sizes, SAR (*Synthetic Aperture Radar*) satellite images [1] has been used. After presenting the oil spill problem, both the Case-Based Reasoning methodology and the Weighted Voting Summarization algorithm are explained. Then, the developed system is described, ending with the results and conclusions.

2 Oil spill risks supervise

When an oil spill is produced, it is crucial to know the progress of the oil slicks generated. The evolution of those slicks should be overseen or even forecasted so that a prediction of the arrival of oil slicks to certain areas can be done. To obtain a precise prediction the behaviour of the oil slicks must be known. Working in an instable and highly changing environment such as the open sea increases the difficulty of providing good predictions.

Data about the oil slicks, like position, shape and size, must be obtained. The best way to acquire that information is by using satellite images, specially SAR satellite images. With those satellite images it is possible to discriminate between normal sea variability and slicks. Distinguish between oil slicks and similar look-alikes is also essential. There have been different approaches to estimate, analyze and forecast the critical situations after an oil spill. In this paper a CBR system is used to generate precise predictions using both historical data and information obtained from satellites.

3 CBR methodology

Case-Based Reasoning methodology is developed from the evolution of knowledge based systems. Past situations are used by CBR systems as the key elements to solve new problems [2]. The *case* base is the main element in the CBR systems structure. The case base stores all the information used to generate new solutions. That information is organized so that similar elements can be easily recovered to solve new problems.

CBR systems learn from past experiences. The process of acquiring new knowledge from past data is divided in four main phases: *retrieval*, *reuse*, *revision* and *retention*. In the *retrieve* phase the most similar cases to the proposed problem are recovered from the case base. With those cases, the system must create a solution adapting them to the new problem, *reusing* them. Once the solution is generated by the system, it must be *revised* to check its correction. If the proposed solution is accepted then it can be eventually *retained* by the system, if there are not redundant information stored, and could serve as a solution to future problems.

4 WeVoS: Weighted Voting Summarization of SOM ensembles

Case-Based Reasoning systems are highly dependent on stored information. The new algorithm presented here, *Weighted Voting Summarization of Som ensembles (WeVoS)* is used to organize the data that is accumulated in the case base. It is also used to recover the most similar cases to the proposed problem.

The main objective of the new fusion of an ensemble of topology preserving maps [3] algorithm presented here, WeVoS, is to generate a final map processed unit by

unit. Instead of trying to obtain the best position for the units of a single map trained over a single dataset, it aims to generate several maps over different parts of the dataset. Then, it obtains a final summarized map by calculating by consensus which is the best set of characteristics vector for each unit position in the map. To do this calculation, first this meta-algorithm must obtain the “quality” [4] of every unit that composes each map, so that it can relay in some kind of informed resolution for the fusion of neurons.

The final map obtained is generated unit by unit. The units of the final map are first initialized by determining their centroids in the same position of the map grid in each of the trained maps. Afterwards, the final position of that unit is recalculated using data related with the unit in that same position in every of the maps of the ensemble. For each unit, a sort of voting process is carried out as shown in Eq. 1:

$$V(p, m) = \frac{|x_{p,m}|}{\sum_1^M |x_p|} \cdot \frac{q_{p,m}}{\sum_1^M q_p} \quad (1)$$

The final map is fed with the weights of the units as it is done with data inputs during the training phase of a SOM, considering the “homologous” unit in the final map as the BMU. The weights of the final unit will be updated towards the weights of the composing unit. The difference of the updating performed for each “homologous” unit in the composing maps depends on the quality measure calculated for each unit. The higher quality (or the lowest error) of the unit of the composing map, the stronger the unit of the summary map will be updated towards the weights of that neuron. The summarization algorithm will consider the weights of a composing unit “more suitable” to be the weights of the unit in the final map according to both the number of inputs recognized and the quality of adaptation of the unit (Eq. 1). With this new approach it is expected to obtain more faithful maps to the inner structure of the dataset.

5 WeVos-CBR

There have already been CBR systems created to solve maritime problems [5] in which different maritime variables have been used. In this occasion, the data used have been previously collected from different observations from satellites, and then pre-processed, and structured to create the case base. The created cases are the main elements to obtain the correct solutions to future problems, through the CBR system. The developed system determines the probability of finding oil slicks in a certain area after an oil spill has been produced.

5.1 Pre-processing and retrieval

When the case base is created the WeVoS algorithm is used to structure it. The graphical capabilities of this novel algorithm are used in this occasion to create a model that represents the actual variability of the parameters stored in the cases. At the same time, the inner structure of the case base will make it easier to recover the most similar cases to the problem cases introduced in the system.

The WeVos algorithm is also used to recover the most similar cases to the problem introduced in the system. That process is performed once the case base is structured keeping the original distribution of the available variables.

5.3 Reuse

After recovering the most similar cases to the problem from the case base, those cases are used to obtain a solution. *Growing RBF networks* [6] are used to generate the predicted solution corresponding to the proposed problem. The selected cases are used to train the GRBF network. This adaptation of the RBF network lets the system grow during the training phase in a gradual way increasing the number of elements (prototypes) which work as the centres of the radial basis functions. The error definition for every pattern is shown below:

$$\epsilon_i = l/p \cdot \sum_{k=1}^p \|t_{ik} - y_{ik}\|, \quad (2)$$

Where t_{ik} is the desired value of the k^{th} output unit of the i^{th} training pattern, y_{ik} the actual values of the k^{th} output unit of the i^{th} training pattern. After the creation of the GRBF network, it is used to generate the solution to the introduced problem. The solution will be the output of the network using as input data the retrieved cases.

5.4 Revision and Retain

In order to verify the precision of the proposed solution, *Explanations* are used [7]. To justify and validate the given solution, the retrieved cases are used once again. The selected cases have their own future associated situation. Considering the case and its solution as two vectors, a distance between them can be measured by calculating the evolution of the situation in the considered conditions. If the distance between the proposed problem and the solution given smaller than the distances obtained from the selected cases, then the proposed solution considered as a good one.

Once the proposed prediction is accepted, it can be stored in the case base in order to serve to solve new problems. It will be used equally than the historical data previously stored in the case base. The *WeVoS* algorithm is used again to introduce new elements in the case base.

6 Results

The WeVoS-CBR system has been checked with a subset of the available data that has not been previously used in the training phase. The predicted situation was contrasted with the actual future situation as it was known (historical data was used to train the system and also to test its correction). The proposed solution was, in most of the variables, close to 90% of accuracy.

Table 1. Percentage of good predictions obtained with different techniques.

<i>Number of cases</i>	RBF	CBR	RBF + CBR	WeVoS-CBR
500	46 %	41 %	44 %	47 %
1000	49 %	46 %	55 %	63 %
3000	58 %	57 %	66 %	79 %
5000	60 %	62 %	73 %	88 %

Table 1 shows a summary of the obtained results. In this table different techniques are compared. The evolution of the results is shown along with the augmentation of the number of cases stored in the case base. All the techniques analyzed improve their results at the same time the number of stored cases is increased. The solution proposed do not generate a trajectory, but a series of probabilities in different areas, what is far more similar to the real behaviour of the oil slicks.

References

- [1] Palenzuela, J.M.T., Vilas, L.G. and Cuadrado, M.S. (2006) Use of ASAR images to study the evolution of the Prestige oil spill off the Galician coast, *International Journal of Remote Sensing*, 27 (10), 1931-1950.
- [2] Aamodt, A. (1991) A Knowledge-Intensive, Integrated Approach to Problem Solving and Sustained Learning, *Knowledge Engineering and Image Processing Group. University of Trondheim*.
- [3] Kohonen, T. (1998) The self-organizing map, *Neurocomputing*, 21 (1-3), 1-6.
- [4] Pözlbauer, G. (2004) Survey and Comparison of Quality Measures for Self-Organizing Maps. In Rauber, J.P.a.G.P.a.A., *Fifth Workshop on Data Analysis (WDA'04)*. Elfa Academic Press, 67--82.
- [5] Corchado, J.M. and Fdez-Riverola, F. (2004) FSfRT: Forecasting System for Red Tides, *Applied Intelligence*, 21, 251-264.
- [6] Karayiannis, N.B. and Mi, G.W. (1997) Growing radial basis neural networks: merging supervised and unsupervised learning with network growth techniques, *Neural Networks, IEEE Transactions on*, 8 (6), 1492-1506.
- [7] Sørmo, F., Cassens, J. and Aamodt, A. (2005) Explanation in Case-Based Reasoning—Perspectives and Goals, *Artificial Intelligence Review*, 24 (2), 109-143.