Information Sciences 180 (2010) 2029–2043

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

A forecasting solution to the oil spill problem based on a hybrid intelligent system

Bruno Baruque [a], Emilio Corchado [b], Aitor Mata [b], Juan M. Corchado [b,*]

[a] Civil Engineering Department, University of Burgos, C/Francisco de Vitoria s/n, 09006 Burgos, Spain
[b] Departamento de Informática y Automática, University of Salamanca, Plaza de la Merced s/n, 37008 Salamanca, Spain

ARTICLE INFO

Article history:
Received 9 June 2009
Received in revised form 5 November 2009
Accepted 28 December 2009

Keywords:
Case-Based Reasoning
Oil spill
Self organizing memory
Radial Basis Function
Ensembles
Fusion algorithms

ABSTRACT

Oil spills represent one of the most destructive environmental disasters. Predicting the possibility of finding oil slicks in a certain area after an oil spill can be critical in reducing environmental risks. The system presented here uses the Case-Based Reasoning (CBR) methodology to forecast the presence or absence of oil slicks in certain open sea areas after an oil spill. CBR is a computational methodology designed to generate solutions to certain problems by analysing previous solutions given to previously solved problems. The proposed CBR system includes a novel network for data classification and retrieval. This type of network, which is constructed by using an algorithm to summarize the results of an ensemble of Self-Organizing Maps, is explained and analysed in the present study. The Weighted Voting Superposition (WeVoS) algorithm mainly aims to achieve the best topographically ordered representation of a dataset in the map. This study shows how the proposed system, called WeVoS-CBR, uses information such as salinity, temperature, pressure, number and area of the slicks, obtained from various satellites to accurately predict the presence of oil slicks in the north-west of the Galician coast, using historical data.

1. Introduction

When an oil spill occurs, the natural risks are evident. Complicated decisions must be made in order to keep the risk from actually becoming a natural disaster. The ability to predict if an area is going to be affected by the slicks generated after an oil spill will be highly useful in making those decisions.

The ocean is a highly variable environment where accurate predictions are difficult to achieve. The complexity of the modelling system increases if external elements are introduced into the analysis. In this case, oil spill data is added to the inherent complexity of the ocean, generating a rough set of elements. To model an environment similar to what is obtained after adding oceanic variables, weather conditions and oil spills, it is necessary to measure different parameters such as wind, current, and pressure. To predict the presence or absence of oil spills in a certain area, their previous positions must be known. That knowledge is provided by the analysis of satellite images, from which the position and size of the slicks are obtained.

The main objective of this interdisciplinary study is to present a new predicting hybrid intelligent model based on the Case-Based Reasoning methodology [58], which accomplishes the different phases of the CBR cycle by using different artificial intelligence techniques. A new algorithm called *WeVoS-SOM* (Weighted Voting Summarization of Self-Organizing Maps) [4] is introduced to organize the structure of the case base and support the recovery and retention of cases. This

* Corresponding author.
E-mail addresses: bbaruque@ubu.es (B. Baruque), escorchado@ubu.es (E. Corchado), aitor@usal.es (A. Mata), corchado@usal.es (J.M. Corchado).

0020-0255/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ins.2009.12.032

novel algorithm improves the results obtained by previous methods [17]. The reuse phase is accomplished by applying the Growing RBF Network (GRBF) [30], a version of the classic Radial Basis Function neural network, which adapts itself to the data and grows as the case base grows. In the revision phase, explanations, which are an automatic way of justifying the solutions generated by the system, are used to internally justify the solution proposed. The *WeVoS-CBR* system presented here has been applied specifically to the data obtained after the Prestige accident in the north-west of Spain. However, the generalization capabilities of the CBR systems will permit the *WeVoS-CBR* system to be applied to other geographical areas and other natural phenomena by slightly adapting the system, which is something quite difficult to achieve with the existing approximations to this kind of problems.

The development of this system was possible thanks to a project founded by the Spanish Ministry of Science and Technology after the Prestige accident. Most of the data used to develop the proposed system was acquired from the ECCO (*Estimating the Circulation and Climate of the Ocean*) consortium [37]. Data related to the oil slicks, including their position and size, were obtained by studying SAR (*Synthetic Aperture Radar*) satellite images [39].

CBR systems have the ability to learn from past situations, and to generate solutions to new problems based on past solutions given to past problems. The system presented in this study combines the efficiency of the CBR systems with other artificial intelligence techniques in order to improve the results and to better generalize from past data. The generalization capabilities of the CBR systems permit this new system to extend the results obtained in the analyzed area to other open ocean areas. This is quite an innovative approach that improves the classic mathematical models, which are incapable of generating possibilities, by applying them to specific local regions. Several forecasting models have been applied to specific geographical zones, where the oceanic behaviour is quite unusual [42], but there does not exist a generic model, such as the one presented in this study, that can be applied in any open ocean region. Hybrid models can use both data and knowledge to forecast trajectories and evaluate possible risks after an oil spill [29].

The hybrid intelligent system proposed in this research incorporates a new ensemble summarization algorithm. The *WeVoS-SOM* algorithm [4] performs the classifications tasks in the CBR structure when creating the case base. This algorithm creates an inner structure within the case base that makes it easier to recover the cases by grouping similar cases together. When a new problem must be solved and the similar cases should be retrieved from the case base, it is very important to recover those cases quickly and accurately. An internal structure such as the one generated by the *WeVoS-SOM* model is crucial in this kind of system. When the similar cases are stored close one to another, then the recovery process does not need to search the entire case base, but only those elements close to the required one, which implies a great reduction of time.

The following section explains the oil spill problem, as well as some previous solutions and models used to try to solve that problem. The Section 3 provices a brief explanation of the CBR methodology detailing the different phases of the CBR cycle and some current applications of the CBR methodology. In Section 4, the *WeVoS* algorithm is developed, including its SOM foundation characteristics [31]. Finally, the hybrid system developed in this study is described, paying special attention to the different techniques used in the four main CBR phases, followed by the experimental results, conclusions and future work.

2. The oil spill problem and existing solutions

Once an oil spill occurs, the progression of the resulting oil slicks must be supervised or even predicted, in order to either determine if an area is going to be contaminated or, better yet, avoid contamination altogether in some critical areas. To get an accurate prediction, it is necessary to know how the oil slicks behave or, at the very least, what the probability is of finding oil slicks in an area. Increasing the number of variables involved in the analysis of the situation, however, also increases the difficulty of obtaining an accurate prediction.

First, the position, shape and size of the oil slicks must be identified. The most precise way to acquire that information is by using satellite images. SAR images are the most commonly used for automatically detecting these slicks [51]. The satellite images show certain areas where there seem to be no waves; but where there are, in fact, oil slicks. Fig. 1 shows an example of a SAR image with oil spills, while Fig. 2 shows the system's interpretation of the image shown in Fig. 1 properly locating the slicks and differentiating them from the coastal areas or even the islands present in the image. The interpretation of the images in conjunction with the variables taken into account (bottom pressure, salinity, wind, current,...) allows the WeVoS-CBR system to generate predictions about the future state of the oil slicks in a particular area. The images recognize the areas to analyze, and then, using the current parameters, a *case* is created. That case is used as a *problem* to be solved, by applying the WeVoS-CBR system.

With the SAR images, it is possible to distinguish between normal sea variability and slicks. It is also important to distinguish between oil slicks and look-alikes [55]. Oil slicks are quite similar to quiet sea areas. If there is not enough wind, the difference between the calm sea and the surface of a slick is less evident, which may result in more mistakes when trying to distinguish between an oil slick and something that is not a slick. This is a crucial aspect in this problem that can also be automatically addressed by a series of computational tools, whereby other meteorological factors are analyzed at the same time as the images, in order to completely distinguish between oil slicks and calm open ocean areas [39].

Once the slicks are identified, it is also essential to know the atmospheric and maritime situation that is affecting the slick at the moment that it is being analysed. Information collected from the satellites is used to obtain the necessary atmospheric