Knowl Inf Syst
DOI 10.1007/s10115-010-0368-y

REGULAR PAPER

Combining case-based reasoning systems and support vector regression to evaluate the atmosphere–ocean interaction

Juan F. De Paz · Javier Bajo · Angélica González · Sara Rodríguez · Juan M. Corchado

Received: 20 March 2010 / Revised: 13 September 2010 / Accepted: 26 November 2010

Abstract This work presents a system for automatically evaluating the interaction that exists between the atmosphere and the ocean's surface. Monitoring and evaluating the ocean's carbon exchange process is a function that requires working with a great amount of data: satellite images and in situ vessel's data. The system presented in this study focuses on computational intelligence. The study presents an intelligent system based on the use of case-based reasoning (CBR) systems and offers a distributed model for such an interaction. Moreover, the system takes into account the fact that the working environment is dynamic and therefore it requires autonomous models that evolve over time. In order to resolve this problem, an intelligent environment has been developed, based on the use of CBR systems, which are capable of handling several goals, by constructing plans from the data obtained through satellite images and research vessels, acquiring knowledge and adapting to environmental changes. The artificial intelligence system has been successfully tested in the North Atlantic Ocean, and the results obtained will be presented in this study.

J. F. De Paz · A. González · S. Rodríguez · J. M. Corchado
Department of Computer Science and Automation, University of Salamanca,
Plaza de la Merced s/n, 37008 Salamanca, Spain

A. González
e-mail: angelica@usal.es

S. Rodríguez
e-mail: srg@usal.es

J. M. Corchado
e-mail: corchado@usal.es

J. Bajo
Faculty of Computer Sciences, Pontifical University of Salamanca,
Compañía 5, 37002 Salamanca, Spain
e-mail: jbajope@upsa.es

J. F. De Paz (✉)
Faculty of Computer Sciences, University of Salamanca, Plaza de la Merced s/n,
37008 Salamanca, Spain
e-mail: fcofds@usal.es

Published online: 14 December 2010

Keywords Case-based reasoning · Support vector regression · Atmosphere–ocean interaction · Decision support

1 Introduction

One of the factors of greatest concern in climatic behavior is the quantity of carbon dioxide (CO_2) present in the atmosphere. Carbon dioxide is one of the greenhouse gases that helps to make the earth's temperature habitable, so long it maintains certain levels [68]. Traditionally, it has been considered that the main system regulating carbon dioxide in the atmosphere is the photosynthesis and respiration of plants. However, thanks to tele-detection techniques, it has been shown that the ocean plays a highly important role in the regulation of carbon quantities, the full significance of which still needs to be determined [71]. Current technology allows us to obtain data and make calculations that were unimaginable some time ago. These data give us an insight into carbon dioxide's original source, its decrease, and the causes for this decrease [44], which allow us to make predictions on its behavior in the future.

CO_2 exchanged between the ocean and the atmosphere is one of the factors that has a high impact on the presence of carbon dioxide in the atmosphere. The ocean's surface and the atmosphere interact and exchange CO_2, which can be absorbed or expelled by the ocean. The conditions for this interaction can be a huge source of knowledge in understanding the carbon cycle and making predictions. Tele-detection techniques provide a huge amount of data about oceanic and atmospheric behavior and requires the use of new computational techniques in order to extract information. The purpose of this study is to provide a hybrid intelligent system based on artificial intelligence techniques that allows the analysis of information contained in the satellite images, from which automatic predictions for decision support [48] in diagnoses of CO_2 exchange can be made.

The system proposed in this work focuses on the detection of the CO_2 exchange rate and is constructed from a case-based reasoning (CBR) system that provides a classification technique based on previous experiences. The CBR paradigm is a type of reasoning that uses past experiences to resolve new problems and is very appropriate for use in scenarios where adaptation and learning abilities are necessary. The CBR system developed in this study receives data from the analysis of satellite images and is responsible for obtaining the CO_2 exchange rate. The hybrid intelligent system proposed in this research integrates CBR and support vector regression (SVR) characterized for their efficiency for data processing and knowledge extraction. In order to acquire intelligent behaviors, it is necessary to provide the monitoring systems with learning capabilities. One of the possibilities is learning from past experiences, which can facilitate cognitive knowledge. The CBR paradigm is aimed at providing learning and adaptation capacities [39–42]. The use of past experiences allows these systems to resolve new problems [42,45]. As SVR is a variation of support vector machines, able to provide regression models for nonlinear datasets, the combination of CBR and SVR provides an added value to the prediction of the CO_2 exchange. This proposal is a step in this direction and the first step toward the development of predictive models based on nonlinear data. The model presented within this work provides great capacities for learning and adaptation to the characteristics of the problem in consideration by using novel algorithms in each of the stages of the CBR cycle that can be easily configured and combined. It also provides results that notably improve those provided by the existing methods for CO_2 analysis [63].

The mission of the intelligent environment presented in this work is to globally monitor the interaction between the ocean's surface and the atmosphere, facilitating the work of oceanographers. Initially, the system is being used in order to evaluate and predict the amount of carbon dioxide (CO_2) absorbed or expelled by the North Atlantic Ocean [6,7,16]. The main purpose of this work is to obtain an architecture that enables the construction of open, distributed, and dynamic systems capable of growing in dimension and of adapting their knowledge according to different changes that take place in their environment.

In the next section, we present the problem that motivates this research. Then, in Sect. 3, the related work is presented. Section 4 describes the SVM and SVR techniques, and Sect. 5 analyzes the CBR paradigm. In Sect. 6, we will describe the approach proposed in this research, together with a case study. Finally, in Sect. 7, some preliminary results and the conclusions will be presented.

2 Air–sea interaction problem

The oceans contain approximately 50 times more CO_2 in dissolved forms than the atmosphere, while the land biosphere including the biota and soil carbon contains about 3 times as much carbon (in CO_2 form) as the atmosphere [71]. The CO_2 concentration in the atmosphere is governed primarily by the exchange of CO_2 with these two dynamic reservoirs. Since the beginning of the industrial era, about 2,000 billion tons of carbon have been released into the atmosphere as CO_2 from various industrial sources including fossil fuel combustion and cement production. It is important, therefore, to fully understand the nature of the physical, chemical, and biological processes, which govern the oceanic sink/source conditions for atmospheric CO_2 [44,71].

The need to quantify the carbon dioxide valence, and the exchange rate between the oceanic water surface and the atmosphere, has motivated us to develop the distributed system, presented here, that incorporates a CBR model capable of estimating such values using accumulated knowledge and updated information. The CBR model receives data from satellites, oceanographic databases, and oceanic and commercial vessels. This information can be used in order to predict the air–sea fluxes of carbon dioxide. The CBR system incorporated is able to optimize tasks such as the interpretation of images using various strategies [62]. The information received is composed of satellite images of the ocean's surface, wind direction and strength, and other parameters such as water temperature, salinity, and fluorescence as can be seen in Fig. 1. An improvement in the monitoring and forecasting methods presented in [5,6,16] has been incorporated in the CBR model presented in this paper.

Information from images can be used in a very proactive way in different scenarios [32,50]. In the carbon dioxide exchange problem, the parameters obtained from the satellite images, which have most influence within our models are air and water temperature, water salinity, wind strength, wind direction and biological parameters such as chlorophyll. These parameters allow us to calculate the variables that define our models, such as the velocity of gas transfer, solubility, or the differentiation between partial pressures on the atmosphere and sea surface. The majority of CO_2 is dissolved in the seawater because of phytoplankton or accumulates of organic materials at the bottom of the ocean. The phytoplankton present in deep areas of the ocean is taken to the surface by surges or surface appearances that are no more than large upward movements of cold water that bring nutrients to the sea surface. The principal cause of these surges are the winds. The way to detect the winds through satellites is to study the images captured by sensors that are sensitive thermal infrared wavelengths (capable of detecting the sea surface temperature (SST)) and to identify cold waters. Another possible way to detect them is to monitor the activity of chlorophyll through sensors within the visible light spectrum range (found between blue and green), which are associated with

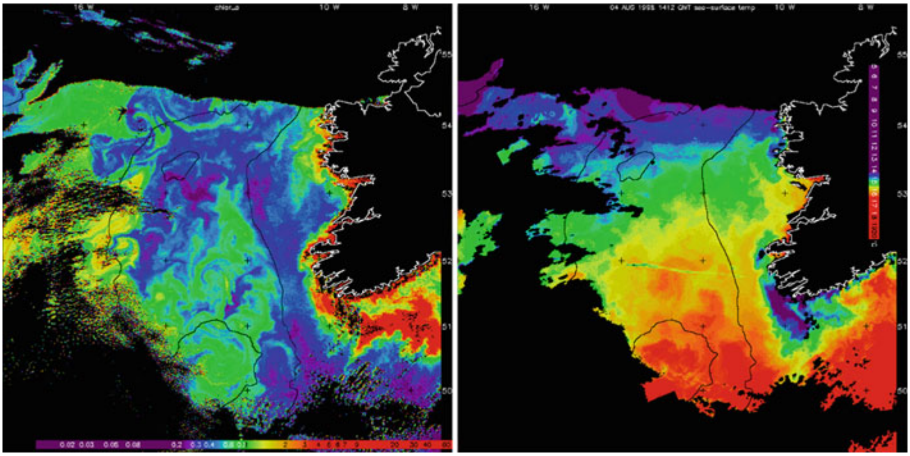

Fig. 1 Satellite color pictures. This images provide information about the Ocean situation

the presence of phytoplankton. In order to obtain the satellite images that contain information about these parameters, it is necessary to use different sensors. The earth observation satellites that have been used to obtain images in the Northern Atlantic are NOAA, Orbview-2, and above all the ENVISAT satellite of the European Space Agency. Below, we shall briefly describe the sensor used in each one of these and the software for the digital processing of the images.

The thermal sensors allow us to measure the surface temperature of the sea. The NOAA satellites are equipped with the Advanced Very High Resolution Radiometer (AVHRR) sensor that is capable of detecting electromagnetic energy reflected by objects present on the earth within five spectrum ranges (three bands in the visible and two in the thermal range). It has a receiving cycle of 12 h with which it is possible to obtain up to six images per day at a resolution of one km^2. In order to determine the SST, the NOAA uses a multi-channel algorithm for the water surface [43]. The ENVISAT satellite has an Advanced Along-Track Scanning Radiometer (AATSR) sensor with which it is capable of exploring the ocean's surface at various infrared and visible frequencies in order to measure the exact temperature. Specifically, the temperature of the sea's surface can be calculated with an accuracy of 0.3°C [72].

There are also sensors that allow us to measure the concentration of chlorophyll. The earth observation satellite Orbview-2 uses a Sea-Viewing Wide Field-of-view Sensor (SeaWIFS) [43,80], which is capable of creating images with information on eight bands or ranges of the electromagnetic spectrum. Of these eight bands, four around the blue–green are used for the detection of chlorophyll. In order to calculate these quantities, the Ocean Chlorophyll 4-band OCTS is used and is included in the SeaWIFS Data Analysis System (SeaDAS) software developed by NASA [60]. The ENVISAT satellite has a Medium Resolution Image Spectrometer (MERIS) with which it is possible to take images of the planet's surface and also the clouds, capturing the light of the visible areas and the infrared of the electromagnetic spectrum. In this way, it is capable of ascertaining the exact color of the ocean's surface and of coastal areas, making it possible to reflect biological activity, to monitor cloud cover, and to detect the movement of vapor from the ocean into the atmosphere. [3,4,72]

The processing of the images obtained may vary depending on the sensor that has taken them [21,22]. The processing of the images is carried out at the CAXIS centre at the

Plymouth Marine Laboratory (PML) [31]. The processing of the thermal images is carried out initially by taking a reading of the images in their original format, as they were received. Then, a calibration and a radiometric correction are made in order to reduce the atmospheric effects, and a reference is made to a known cartography base. The next step is to mask the clouds and the land in order to eliminate distortions. Lastly, the sea surface temperature is calculated applying a suitable algorithm. In order to process images of the chlorophyll concentration, a reading is made of the images and decoded when necessary. Meteorological and ozone files are requested by the software. The clouds and land are masked and the chlorophyll image is calculated. Lastly, reference is made to a known cartographic base and compositions and mid-point images are made, which can take some days.

The system presented aims at modeling the flux of carbon dioxide exchanged between the atmosphere and the ocean surface. The oceans contain approximately 50 times more carbon dioxide in dissolved forms than the atmosphere, while the land biosphere including the biota and soil carbon contains about 3 times as much carbon (in carbon dioxide form) as the atmosphere [71]. The carbon dioxide concentration in the atmosphere is governed primarily by the exchange of carbon dioxide with these two dynamic reservoirs. Since the beginning of the industrial era, about 2,000 billion tons of carbon have been released into the atmosphere as carbon dioxide from various industrial sources including fossil fuel combustion and cement production. This amount, which is about 35% of the total amount of carbon in the pre-industrial level, corresponds to approximately 590 billion tons as carbon. At present, atmospheric carbon dioxide content is increasing at an annual rate of about 3 billion tons, which corresponds to one half of the annual emission rate of approximately 6 billion tons from fossil fuel combustion. Whether the missing carbon dioxide is mainly absorbed by the oceans or by the land and their ecosystems has been debated extensively over the past decade.

It is important, therefore, to fully understand the nature of the physical, chemical, and biological processes that govern the oceanic sink/source conditions for atmospheric carbon dioxide [44,71]. Satellite-borne instruments provide high-precision, high-resolution data on atmosphere, ocean boundary layer properties, and ocean biogeochemical variables, daily, globally, and in the long term. All these new sources of information have changed our approach to oceanography, and the data generated need to be fully exploited. Wind stress, wave breaking, and the damping of turbulence and ripples by surface slicks all affect the air–sea exchange of carbon dioxide. These processes are closely linked to the "roughness" of the sea surface, which can be measured by satellite radars and microwave radiometers. Sea surface roughness consists of a hierarchy of smaller waves upon larger waves. Different sensors give subtly different measurements of this roughness.

Our final aim is to model both the open ocean and shelf seas, and it is believed that by assimilating Earth Observation (EO) data into artificial intelligence models, these problems may be solved. Earth observation data (both for assimilation and for validation) are vital for the successful development of reliable models that can describe the complex physical and biogeochemical interactions involved in marine carbon cycling. Satellite information is vital for the construction of oceanographic models and in this case to produce estimates of air–sea fluxes of carbon dioxide with much higher spatial and temporal resolution, using artificial intelligence models than can be achieved realistically by direct in situ sampling of upper ocean carbon dioxide. To handle all the potentially useful data to create daily models in a reasonable time and with a reasonable cost, it is necessary to use automated distributed systems capable of incorporating new knowledge. Our proposal is presented in the following section.

3 Related work

It is possible to find different systems in literature aimed at predicting CO_2 exchange rates [36,37,67]. These works propose an approach based on obtaining regression models that are generated manually by experts. The works presented in [36,37] focus on the variation of the exchange of CO_2 produced during the day and during the night, while the work presented in [67] prioritizes the difference of pressures that exists between the ocean surface and the air. The regression models proposed in these works have, in general, a high level of complexity and sometimes require the incorporation of new variables once the model has been generated, which means recalculating the equations of the model. In this sense, the estimation of the CO_2 exchange rate obtained by means of manual models presents deficiencies when working in dynamic environments, where the system needs to automatically adapt itself to the changes that occur in its surroundings and evolve over time.

An alternative to the manual approaches are traditional prediction systems, but these systems require an assisted process to design the models, which introduces a high level of complexity in the development process. The most prevalent alternatives are those oriented to develop prediction systems based on artificial neural networks technology [27]. An example of these kind of systems can be found in [16,30], where agents use the CoHeL IBR system to achieve their goals [16]. The Cooperative Maximum Likelihood Hebbian Learning (CoHeL) method is a novel approach that features selection, in which the aim is to visualize and extract information from complex and highly dynamic data. The model proposed is a mixture of factor analysis and exploratory projection pursuit based on a family of cost functions proposed by Fyfe and Corchado [27], which maximizes the likelihood of identifying a specific distribution in the data while minimizing the effect of outliers. It employs cooperative lateral connections derived from the Rectified Gaussian Distribution [69] in order to enforce a more sparse representation in each weight vector. This method is used for the clustering of instances, and during the retrieval stage of the IBR cycle, the adaptation step is carried out using a radial basis function network. Finally, the system is updated continuously with data obtained from the CaStore agents. The CoHeL IBR system is described in [16]. Some similar works using artificial neural networks can be found in problems of similar characteristics, specifically those related to time series. Thus, it is possible to find neural network-based models to predict energy consumption [29,34] or gas consumption [55].

However, the use of artificial neural networks or genetic algorithms to resolve this kind of problems presents certain problems during the training stage of the artificial neural network or genetic algorithm. The problems are caused because it is necessary to have available a minimum of data to successfully complete the training stage, and not always it is possible to have such amount of data available.

Other approaches to the artificial neural networks are the automatic regression models. In fact, most of the existing proposals aimed at resolving the CO_2 exchange rate prediction problem propose the use of regression models, as can be seen in [36,37,67]. The main reason to use regression models, as indicated in [52], is that the CO_2 exchange cannot be studied as a linear model. The existing regression models are mainly based on the use of support vector machine for Regression (SVR) [70,74,75] as the key technique. The SVR technique has been used in different case studies to make climatic predictions [73] or pollution predictions [61], or more generally, to predict time series [64]. Basically, the objective of SVR is to transform the input data into a high-dimensionality space where the variables can be linearly separated. Based on the linear ability to be separated, it is possible to generate linear models to approximate the data by means of the minimization of an objective function.

Other approaches propose an analysis and design methodology that facilitates the implementation of CBR agent-based distributed artificial intelligent systems. Moreover, the architecture takes into account the fact that the working environment is dynamic and therefore it requires autonomous models that evolve over time. In order to resolve this problem, an intelligent environment has been developed, based on the use of CBR agents, which are capable of handling incorporated goals; constructing plans from the data obtained through satellite images and research vessels, acquiring knowledge and of adapting to environmental changes.

This article proposes a new perspective, where CBR and SVR are combined to provide predictions based on past experiences and regression models for nonlinear datasets. In this sense, the approach is innovative since it proposes step toward the development of predictive models based on nonlinear data. The CBR paradigm has been selected because it provides learning and adaptation capacities, which are very appropriated in dynamic changing environments. Besides, SVR facilitates regression models for nonlinear databases that represents a clear advantage for the prediction of carbon dioxide exchange, as explained in the following section.

4 Support vector regression

SVR comes from support vector machine (SVM) and is specialized in obtaining regression models by means of a change in the dimensionality of the data. The SVM technique is a supervised learning technique that is applied to the classification and regression of different elements. The SVM model is commonly used in different fields, as chemical [47], modeling and simulation [79], data mining [46] or text mining [49], because it facilitates working with data that cannot be adjusted to linear models [74], initially conceived to obtain classifications in linear separable problems, by means of finding a hyperplan able to separate the elements of a set. The SVM technique also allows separation of nonlinear data. To obtain nonlinear separation, SVM performs a mapping of the initial data into a high-dimensionality space, where the data can be linearly separable using specific functions. Given that the dimensionality of the new space can be very high, most of the time it is not viable to use hyperplanes to obtain linear separation. As a solution, nonlinear functions called kernels are used.

The SVR technique is a variation of SVM to generate regressions [70,74,75]. The aim is to adjust the data. As in the case of SVM, there is a mapping of the input data into a high-dimensionality space. In this new space, the regression can be carried out without the initial limitations. Equation (1) shows the linear regression obtained by means of $g_j(x)$ functions that transform the input vectors from their initial coordinates to a high-dimensionality space.

$$f(\vec{x}, \vec{w}) = \sum_{j=1}^{m} w_j g_j(\vec{x}) + b \tag{1}$$

Figure 2 presents an example of linear regression obtained after transforming the input data to be represented in a high-dimensionality space by means of nonlinear transformations. The staple lines represent the interval that delimits the space where non-approximation error is allowed.

As can be observed in Fig. 2, there is an approximation error for the different points and SVR tries to minimize these errors by defining an optimization problem. The first step is to define a loss function, specifically oriented to evaluate the error during the regression process. There exists different functions for the loss function, as a variety of Quadratic

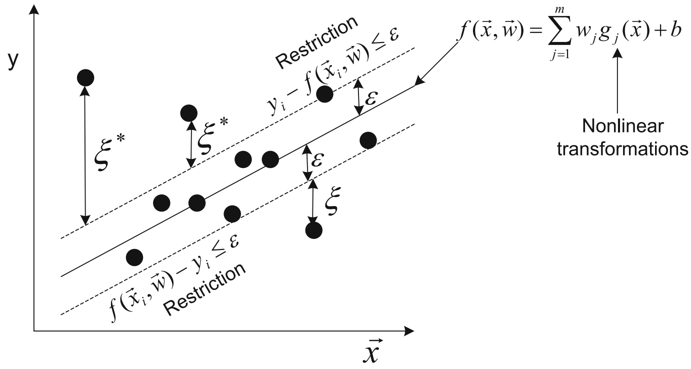

Fig. 2 Results obtained after applying SVR to the initial data. The data is transformed to a high-dimensionality space using nonlinear functions

functions, Huber, ε-insensitive, etc. [70]. It is necessary to define the loss function before stating the optimization problem, and then establish the restrictions. In general, the preferred loss function is the ε-insensitive [70], defined as shown in Eq. (2)

$$L_\varepsilon\left(y, f\left(\vec{x}, \vec{w}\right)\right) = \begin{cases} 0 & \text{if } |y - f(x, w)| \le \varepsilon \\ |y - f(x, w)| - \varepsilon & \text{otherwise} \end{cases} \tag{2}$$

Once the error function has been defined, it is necessary to identify the objective function for the optimization problem. The problem basically consists of finding the most simple vector $\vec{w}$ satisfying certain error restrictions given by (2). To do so, the vector norm is minimized $\|\vec{w}\|^2$, because if the vector $\vec{w}$ does not contain high values, it is possible to minimize the errors. The objective function must be defined by taking into account the vector norm and the condition that indicates that the error allowed should be minor than ε:

$$\begin{aligned} &\min_{w} \frac{1}{2}\|w\|^2 \\ &\text{st} \\ &y_i - f\left(\vec{x}_i, \vec{w}\right) \le \varepsilon \\ &f\left(\vec{x}_i, \vec{w}\right) - y_i \le \varepsilon \end{aligned} \tag{3}$$

The problem taken into consideration can sometimes not be satisfied by the imposed restrictions, that is, it is not possible to find a function f satisfying $y_i - f(\vec{x}_i, \vec{w}) \le \varepsilon$ as can be seen in Fig. 2. As a solution, slack variables are taken into consideration, because they will allow us to find solutions in these cases. Slack variable are the values ξ_i, ξ_i^* and are defined based on the distance shown in Fig. 2. The introduction of these variables into the model implies the modification of the optimization problem presented in (2), as it becomes necessary to modify the objective function to minimize the impact of the addition of the slack variables. It is also necessary to modify the initial restrictions [19,70].

$$\begin{aligned} &\min_{w,\xi_i,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\left(\xi_i - \xi_i^*\right) \\ &\text{st} \\ &y_i - f\left(\vec{x}_i, \vec{w}\right) \le \varepsilon - \xi_i^* \end{aligned}$$

$$f(\vec{x}_i, \vec{w}) - y_i \leq \varepsilon - \xi_i$$
$$\xi_i, \xi_i^* \geq 0, \quad i = 1, \ldots, n \tag{4}$$

The restrictions are defined taking into account the loss function ε-insensitive in such a way that the following condition be satisfied: the estimated values do not have to exceed the value $\varepsilon - \xi_i^*$, as shown in Fig. 2, C represents a constant value.

Finally, to resolve the problem, the dual is obtained [18,24], and the result is given as the solution of the following optimization problem.

$$\begin{aligned} &\max f(\vec{x}) = \sum_{i=1}^{i} \left(\alpha_i - \alpha_i^*\right) k(x_i, x_j) + b \\ &\text{st} \\ &0 \leq \alpha_i^* \leq C \\ &0 \leq \alpha_i \leq C \end{aligned} \tag{5}$$

The transformations $g_j(x)$ presented in Eq. (1) are executed by means of Kernel functions aimed at optimizing the performance of the calculus carried out in the new high-dimensionality space. A variety of predefined kernel functions exist such as Radial Basis kernel, Polynomial kernel, and Linear kernel. These functions provide a transformation of the problems into high-dimensionality spaces with efficient calculus. An example of kernel function is the Polynomial degree d kernel shown in Eq. (6)

$$k(x_i, x_j) = (x_i \cdot x_j + 1)^d \tag{6}$$

By means of SVR, data can be represented in high-dimensionality spaces and can be linearly adjusted. This fact makes SVR very appropriate to be applied to the problem at hand i.e. estimation of the CO_2 exchange rate between the ocean surface and the atmosphere. This is because, as demonstrated in [52], this problem cannot be linearly resolved in the initial space of coordinates. However, it is necessary to note that the input data selected to be passed to SVR is a key factor that highly influences the final result. That is the reason why, in order to look for an estimation as to how the problem can be linearly approximated, the CBR problem is proposed as a reasoning mechanism in this work. The fundamentals of CBR are explained in detail in the next section.

5 Case-based reasoning systems

Most computational systems are algorithmic and work with exact information. However, most applications in the real world require more complex systems, where there is a need for interaction with changing environments and with a certain degree of uncertainty. The CBR paradigm is very appropriate for providing solutions in dynamic environments, where it is usual to have a partial vision of the problem and there is a limitation for the resources available. The purpose of CBR is to solve new problems by adapting solutions that have been used to solve similar problems in the past [1]. The primary concept when working with CBR is the concept of case. A case can be defined as a past experience and is composed of three elements: a problem description that describes the initial problem, a solution that provides the sequence of actions carried out in order to solve the problem, and the final state that describes the state achieved once the solution was applied.

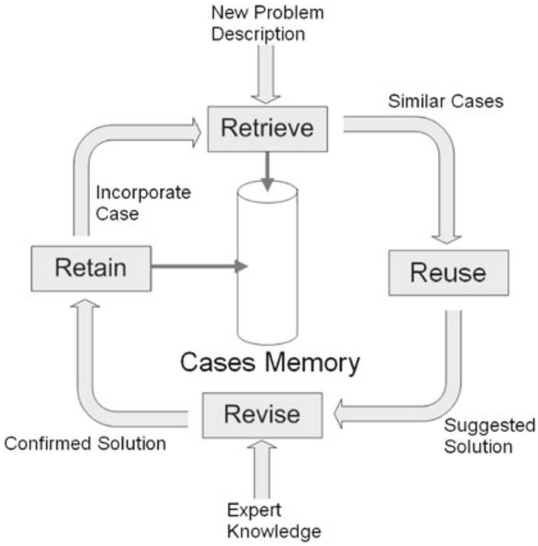

Fig. 3 Diagram of a generic CBR reasoning cycle

Algorithm 1

Case: <Problem, Solution, Result>
Problem: initial_state
Solution: sequence of <action, [intermediate_state]>
Result: final_state

A CBR manages cases (past experiences) to solve new problems. The way cases are managed is known as the CBR cycle and consists of four sequential steps that are recalled every time a problem needs to be solved: retrieve, reuse, revise, and retain. Each of the steps of the CBR life cycle requires a model or method in order to perform its mission. The algorithms selected for the retrieval of cases should be able to search the case base and select the problem and corresponding solution most similar to the new situation. In our case study, the algorithms conducted a filtering of variables, recovered important variables from the cases, and determined which where most influential in the classification process. Once the most important variables have been retrieved, the reuse phase begins and the solutions for the retrieved cases are adapted so that clustering may be obtained. Once this grouping is accomplished, the next step is knowledge extraction. The revise phase consists of an expert revision for the proposed solution, and finally, the retain phase allows the system to learn from the experiences obtained in the three previous phases, consequently updating the cases memory. Figure 3 shows a diagram of the techniques applied in the different stages of the CBR cycle.

The CBR systems are very appropriate for providing solutions for the problem at hand, because of their special ability to learn from past experiences. Monitoring CO_2 exchange is a problem that depends greatly on the conditions of the sea and the atmosphere, and these conditions can be compared to similar past conditions in order to make predictions.

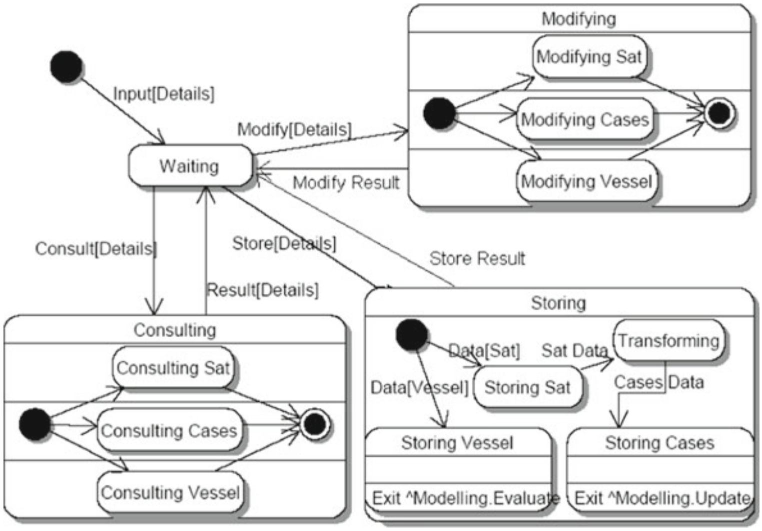

Fig. 4 Diagram of the CBR system's possible states

6 CBR system for monitoring CO_2 exchange rate

The model proposed in this paper presents a new CBR system, which models the air–sea CO_2 exchange rate by means of a novel technique based on a new regression model for nonlinear data. The CBR system has two aims. The first one is to generate models that are capable of predicting the atmospheric–oceanic interaction in a particular area of the ocean in advance. The second one is to permit the use of such models. In Fig. 4, we can see that the system possesses two principal states: one to generate the forecasting models and the other to permit the use of the models. The system can be found in three possible states: a state in which it is awaiting requests; a state in which it is modifying stored data; and, a state in which the system carries out operations to store data. Some of these operations involve obtaining particular parameters that characterize an image. An additional state allows it to be ready to receive new requests when it has no current tasks waiting to be carried out.

Moreover, the reasoning cycle is one of the activities carried out by the system. We can see how the reasoning cycle of a CBR system is included among the activities, composed of stages of retrieval, reuse, revise, and retain. Also, an additional stage that introduces expert's knowledge is used.

Figure 5 shows the internal structure of our CBR system. Problem description (initial state) and solution (situation when final state is achieved) are represented as a set of values related to the oceanic and atmospheric status, the final state is the solution achieved for the problem (the predicted flux of CO_2), and the sequences of actions are the steps carried out in each of the stages of the CBR cycle. The structure of a case for the CO_2 exchange problem can be seen in Table 1.

Table 1 shows the description of a case: DATE, LAT, LONG, SST, S, WS, WD, Fluo_calibrated, SW pCO_2, and Air pCO_2. Flux of CO_2 is the value to be identified (predicted). DATE represents the date of the case, LAT represents the latitude of the location where the data has been obtained and LONG, the longitude in decimal degrees. SST represents the

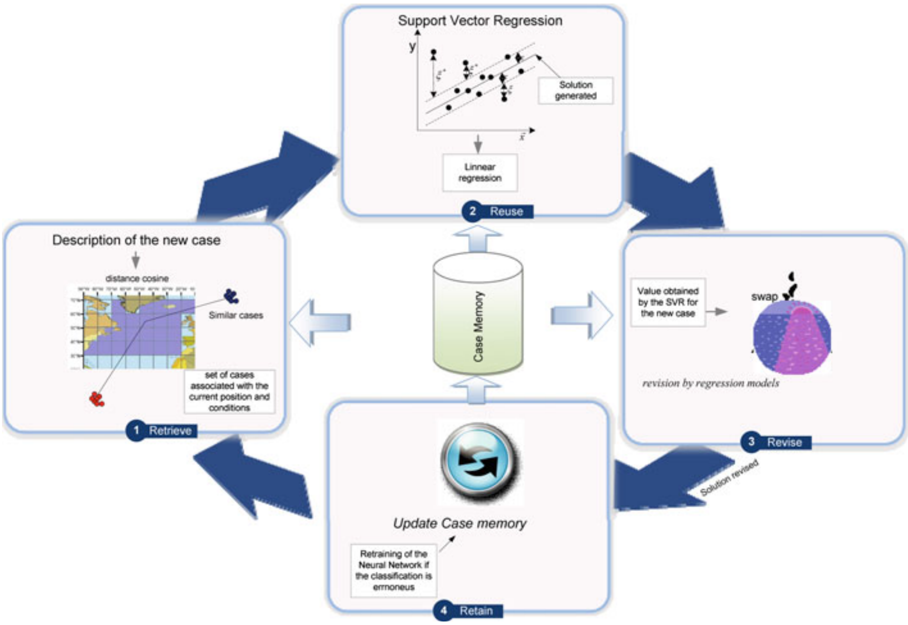

Fig. 5 Internal structure of CBR system proposed in this study

Table 1 Case attributes

Case field	Measurement
DATE	Date (dd/ mm/yyyy)
LAT	Latitude (decimal degrees)
LONG	Longitude (decimal degrees)
SST	Temperature (°C)
S	Salinity (unitless)
WS	Wind strength (m/s)
WD	Wind direction (unitless)
Fluo_calibrated	Fluorescence calibrated with chlorophyll
SW pCO_2	Surface partial pressure of CO_2 (micro atmospheres)
Air pCO_2	Air partial pressure of CO_2 (micro atmospheres)
Flux of CO_2	CO_2 exchange flux $(\text{moles}/\text{m}^2)$

temperature of the ocean and S, the salinity. WS is the wind strength and WD is the wind direction. Fluo_calibrated represents the fluorescence calibrated with chlorophyll.

The CBR cycle is implemented through goals and plans. When the goal corresponding to one of the stages is triggered, different plans (algorithms) can be executed concurrently to achieve the goal. Each plan can trigger new sub-goals and, consequently, cause the execution of new plans.

The problem description represents the state of the problem, with certain knowledge about the surroundings. In our problem, we shall use the attributes DATE, LAT, LONG, SST, S, WS, WD, Fluo_calibrated, SW pCO_2, and Air pCO_2.

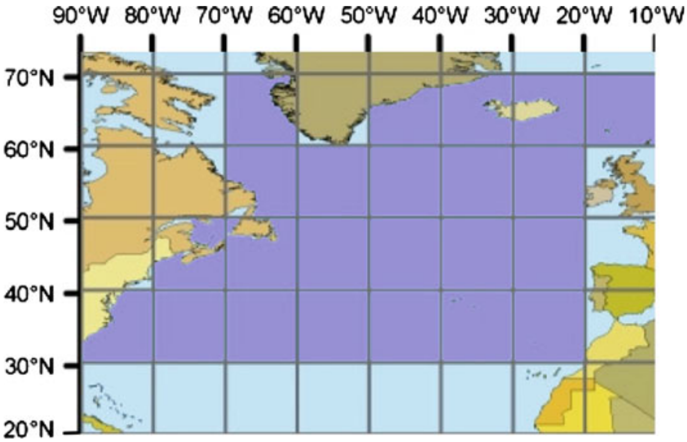

Fig. 6 Oceanic regions of the North Atlantic Ocean taken into consideration for this study

The objectives of the CBR represent those final states that the system wishes to arrive or reach. In this case, it deals with three goals:

- Predict the flux of carbon dioxide exchanged between the sea surface and the atmosphere, using a window of two or 3 weeks.
- Calculate the best parameters to use in order to improve the prediction for different window sizes.
- Calculate the most suitable prediction window in relation to a maximum error percentage allowed.

The sequence of actions to be carried out is generally formed by the stages of the reasoning cycle and the different algorithms executed in each one of those stages. In general, there will be various pre-defined plans or intentions available to the system, which can be called up and modified at the time of execution. The selection of plans is made through the CBR system. The CBR motor is divided into four sequential stages and different algorithms can be used in each one. The reasoning structure is presented in detail in the next paragraphs.

6.1 Retrieve

The prediction for the CO_2 exchange rate is obtained from the parameters shown in Table 1. The prediction is carried out taking into consideration different regions of the Atlantic Ocean and, in order to obtain an effective prediction, the system needs to recover the appropriated past experiences. That is, those cases that contain problem descriptions for similar latitudes and longitudes. In order to establish this first filter in the retrieve stage, the oceanic region taken into consideration for this study was divided into grids of 10° for the latitudes and longitudes, as represented in Fig. 6. Figure 6 shows the regions, in dark color, of the North Atlantic ocean where the tests were carried out. The predictions and estimations are provided for the complete grid as a set.

Once a region has been selected, the selection of the most similar case study is performed according to the cosine distance applied to the following normalized set of variables SST, S, WS, WD, Fluo_calibrated, and Air pCO_2. The cosine distance is used to avoid data normalization and corresponding problems with the data units.

6.2 Reuse

Once the most similar cases have been retrieved, the regression model is generated. As indicated in Sect. 4, the technique that will be used to create the regression model is support vector regression (SVR). The input vector x represents a dataset with the structure presented in Table 1. The input vector can be represented as $\vec{x}$=(DATE, LAT, LONG, SST, S, WS, WD, Fluo_calibrated, SW pCO_2, and Air pCO_2). The regression is obtained making use of all the vectors provided by the most similar cases retrieved in the previous stage of the CBR cycle, and the SVR is calculated following the algorithm presented in Sect. 4. The regression model is used to estimate the swap of the new case, that is the index in the memory of cases, which will be used to generate the prediction value p^*.

6.3 Revise

This phase is performed in an automatic fashion and takes into account the error rate provided by the SVM. The error rate is calculated from the previous existing data using the coefficient of variation, in such a way that if the value obtained is minor than a pre-fixed value, then the prediction can be considered as successful. It is necessary to take into account that once the real data are obtained, the predicted exchange values are eliminated. The estimated values are only used to obtain prediction models under different conditions.

Moreover, during the revision stage, an equation (F) [44] is used to validate the proposed solution p^*.

$$F = k\mathrm{so}(\mathrm{pCO_2SW} - \mathrm{pCO_2AIR}) \tag{7}$$

where F is the flux of CO_2 and k is the gas transfer velocity. Then,

$$k = (-5.204\ \mathrm{Lat} + 0.729\ \mathrm{Long} + 2562.765) \big/ 3600 \tag{8}$$

where Lat is the Latitude, Long is the Longitude, and so is the Solubility. Then, it is verified that:

$$\mathrm{so} = e^{\left(\frac{93.4517}{100\mathrm{tk}} - 60.2409 + 23.3585 \log(100\mathrm{tk}) + \mathrm{s}(0.023517 - 0.023656 \cdot 100\mathrm{tk} + 0.0047036 \cdot 1002\mathrm{tk})\right)} \tag{9}$$

$$\mathrm{tk} = 273.15 + \mathrm{t} \tag{10}$$

$$\mathrm{pCO_2} = \mathrm{A} + \mathrm{BLong} + \mathrm{CLat} + \mathrm{DSST} + \mathrm{EYear} \tag{11}$$

where SST is the temperature of the marine surface or air as it corresponds to pCO_2SW or pCO_2AIR. The coefficients of the equation depend on the month, as shown in Table 2.

During the revision, the system compares the obtained F value with the predicted one, and if the prediction differs in less than 10%, the case is stored on the base of beliefs.

6.4 Retain

This phase provides learning capabilities, since the system is able to learn from past experience. In this phase, the case's information (case description, solution, and efficiency) is stored in the memory of cases and will be available for use as previous knowledge for future predictions. Only those cases considered as efficient in the reuse phase are stored. Once the real exchange value is calculated, the information is updated.

Table 2 Months\coefficients values

Months\coefficients	A	B	C	D	E
Feb	−2,488	−0.42	4.98	−12.23	1.38
May	−7,642	−0.9	−1.74	−20.77	4.14
Jun	−4,873	−0.85	1.3	−15.64	2.66
Jul	−7,013	−0.025	3.66	−7.07	3.64
Aug	−3,160	−0.69	0.84	−11.31	1.8
Sep	−1,297	0.43	−4.19	−17.06	1.05
Oct	83	−0.81	4.81	−10.92	0.076
Nov	747	0.2	−0.73	−17.3	−0.062
Dec	−4,306	0.38	−0.22	−17.13	2.45

7 Experimental results

In this section, we present the experiments developed to evaluate the approach presented in Sect. 6. A first experiment evaluated the influence of the relationship between the latitude and the longitude in the predictions. A second experiment evaluates the predictive capacities of the system using a SVR with 365 existing cases from different grids. A third experiment evaluates the prediction abilities of the system when cases from the current grid are retrieved. Moreover, the results are compared to alternative predictive techniques. The last of the experiments evaluates the behavior of the CBR system taking into consideration the evolution of the CBR system related to the growth of the memory of cases.

The first step was to complete a study on the relation that can be established between the variables of latitude and longitude and how this relationship can affect the CO_2 exchange rate. Figure 7 shows the result obtained for the dataset used for the experiments. As seen in Fig. 7, the exchange rate has a linear dependence with the latitude, while the influence of the longitude is lower. However, it is also possible to appreciate that the influence of the longitude should not be considered as insignificant. This relationship was observed by means of a simple study of the correlation that exist between the variables.

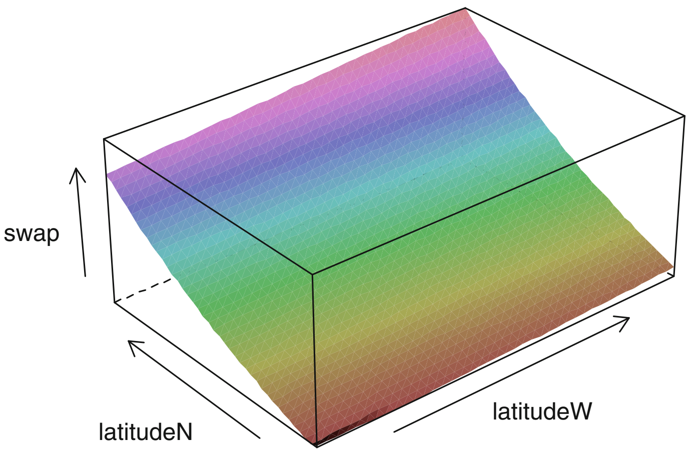

Fig. 7 The relationship of latitude and longitude in the exchange of CO_2

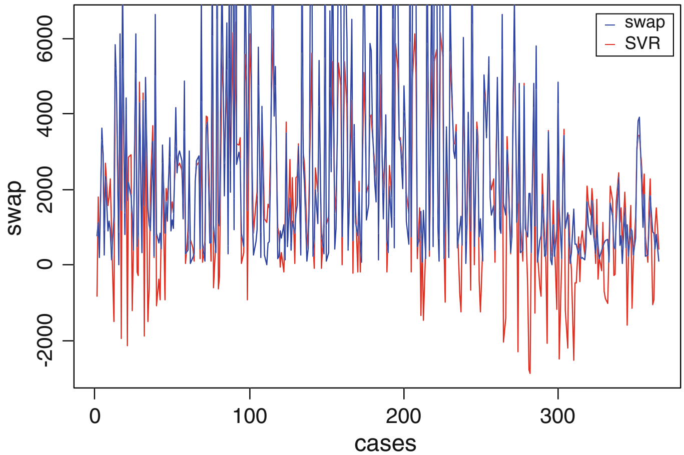

Fig. 8 Predictions obtained when previous similar cases are selected

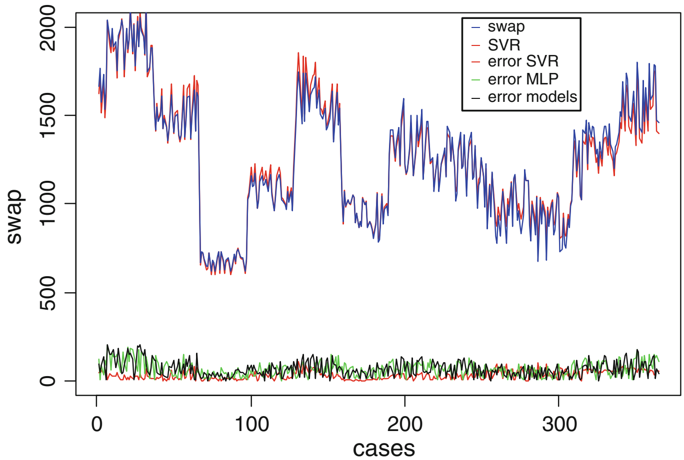

Fig. 9 Comparison between the real values and the prediction values for the CO_2 exchange rate

In order to make evident the need to carry out a separation of the data in latitudes and longitudes, Fig. 8 shows the results obtained after calculating the predictions using SVR with a dataset of 365 cases distributed in a homogeneous manner along the North Atlantic Ocean. The kernel function used for the experiments was polynomial and the loss function was ε-insensitive. The blue lines in Fig. 8 represent the real value of the data, and the red lines represent the predicted values. As can be seen in Fig. 8, the error rate obtained in this experiment is very high compared with the error rate obtained in the experiment presented in Fig. 9. The numerical values represent the millions of tonnes of carbon dioxide that have been absorbed (negative values) or generated (positive values) by the ocean during each of the 3 months.

To evaluate the prediction capacities of the systems presented in this study, different tests were performed along the North Atlantic oceanic region with data obtained during 2009. In each of the tests, when a case containing the description of an oceanic area was

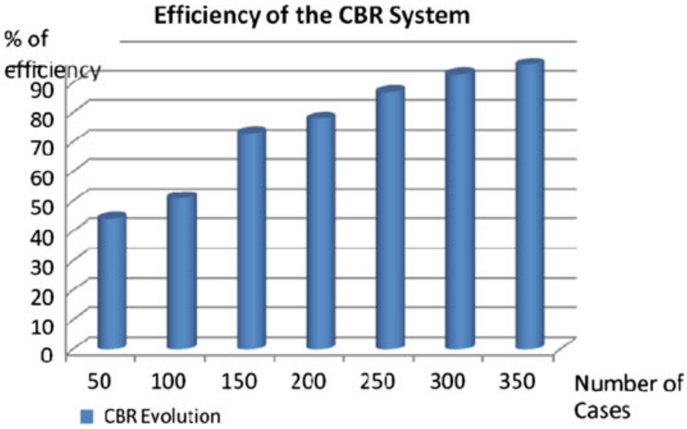

Fig. 10 Evolution of the mean accuracy of the CBR system related to the number or cases

introduced to the system, the most similar cases in the grid with the same latitude and longitude as the new case were taken into consideration to obtain the prediction. Figure 9 shows the results obtained from the experiment. The blue line represents the real value, and the red line represents the predicted value. Moreover, Fig. 9 shows the absolute error rate obtained for the predicted value (red line) provided by the SVR. The mean absolute error rate obtained was 31.43, with an error deviation of 39.63. The mean error percentage obtained was 2.5%.

The absolute error rate obtained with the SVR has been compared to the error rate provided by alternative techniques, such as the multilayer perceptron and the oceanographers' manual models. Figure 8 shows the absolute error rate obtained for each of these predictions. The green line represents the error introduced in the system when the prediction is carried out using a multilayer perceptron. The multilayer perceptron used 27 neurons in the hidden layer, and the final error percentage obtained was 5.1%. Finally, the error rate introduced in the system when the manual models are considered was 6.7%.

Figure 10 shows the evolution of the CBR system related to the number of cases stored in the memory of cases. As can be seen in Fig. 10, as the number of cases increase, the performance of the system improves. This is a typical behavior in CBR systems, because they make use of past experiences to make predictions and they need a certain number of cases to acquire the necessary knowledge. Each of the bars represent the normalized average value of the relationship between the predicted and the real value. Table 3 presents a evaluation of the evolution of the mean accuracy for the SVR and MLP. Moreover, Table 3 shows the mean error obtained for the predictions related to the number of cases used as a base of previous experience.

Oceanographers interaction with the system was smooth and easy. They are evaluating its use in a routine way. The system has been tested during the last 3 months of 2009, and the results have been very accurate, with an error percentage of 2.5. The oceanographers have noted that the CBR system may facilitate their work and provide a highly appreciated decision support tool. They believe that this hybrid neural intelligent architecture has more advantages than disadvantages and that the system has helped them to predict the CO_2 exchange rate. However, they tend to argue that the hybrid neural intelligent architecture should incorporate a shared memory of cases to compare data from different oceanic areas.

Table 3 Evolution of the mean accuracy and mean error related to the number of cases used for the prediction

Cases	SVR		MLP	
	% Mean accuracy	Mean error	% Mean accuracy	Error deviation
50	44.2	730.5	40.3	781.7
100	50.9	634.2	48.4	668.1
150	73.3	348.3	62.3	491.1
200	78.4	281.3	72.2	362.0
250	88.6	148.9	85.8	183.8
300	93.4	85.5	87.7	159.9
350	96.3	32.3	94.3	74.0

8 Conclusions

The application of artificial intelligence techniques [1] is extremely useful in a field like oceanography and specifically in the study of the carbon dioxide exchange between the ocean surface and the atmosphere. The intelligent environment that has been presented in this article allows oceanographers to maintain a seamless, unobtrusive, and often invisible yet controllable interaction with the available technology. The appropriate Store or Vessel can be selected through a simple interface that only presents the necessary information and avoids showing too many elements on the screen. The oceanographers themselves can decide the amount of elements that they wish to see. A request for the creation of a new model can be made (for which it will be necessary to enter the appropriate parameters). An enquiry about the models stored can be made. Also, models can be evaluated by entering real data or saving corresponding data to the models currently being used. The intelligent system presented in this study incorporates advanced reasoning abilities to predict the level of exchange in a particular ocean zone. The system has been compared to the existing alternatives and evaluated in different experiments, and the results obtained demonstrate that the proposed system improves the prediction capacities of the previous approaches and allows predictions for nonlinear datasets.

This study has presented a CBR intelligent system to predict and monitor the CO_2 exchange rate in the North Atlantic Ocean. It applies a hybrid reasoning system specifically designed to analyze data from satellite images and vessels and predict potential CO_2 fluxes in order to provide an innovative method for exploring the CO_2 exchange prediction process and extract knowledge. This knowledge helps human experts to understand the prediction process and to obtain conclusions about the relevance of the situation of the oceanic environment.

The oceanographers that tested the system have noted that the CBR system may facilitate their work and provide a highly appreciated decision support tool. They believe that this hybrid neural intelligent architecture has more advantages than disadvantages and that the system has helped them to predict the CO_2 exchange rate. However, they tend to argue that the hybrid neural intelligent architecture should incorporate a shared memory of cases to compare data from different oceanic areas.

Future work is focused on improving the retain stage of the CBR system, in order to improve the learning capacity of the approach.

References

1. Aamodt A, Plaza E (1994) Case-based reasoning: foundational issues, methodological variations, and system approaches. AI Commun 7:39–59
2. Aarts E (2004) Ambient intelligence, third international conference, AH 2004. LNCS, pp 1
3. Aiken J, Fishwick JR, Lavender SJ, Barlow R, Moore GF, Sessions H, Bernard S, Ras J, Hardman-Mountford NJ (2007) Validation of MERIS reflectance and chlorophyll during the BENCAL cruise October, 2002: preliminary validation and new products for phytoplankton functional types and photosynthetic parameters. Int J Remote Sens 28(3–4):497–516
4. Aiken J, Hardman-Mountford N, Ufermann S, Woolf D, Challenor P, Robinson I, the CASIX team (2005) Exploiting ENVISAT-ERS data for deriving air-sea fluxes of CO_2. In: Proceedings of the ENVISAT symposium
5. Bajo J, Corchado JM (2005) Evaluation and monitoring of the air-sea interaction using a CBR-agents approach. In: Proceedings of the ICCBR 2005 LNCS, vol 3620, pp 50–62
6. Bajo J, Corchado JM (2005) Multiagent architecture for monitoring the North-Atlantic carbon dioxide exchange rate. In: Proceedings of the CAEPIA 2005 LNCS, vol 4177, pp 321–330
7. Bajo J, Corchado JM, De Paz Y, De Paz JF, Martín Q (2006) A multiagent recommending system for shopping centres. In: Proceedings of the workshop on recommender systems (ECAI 2006), vol 1, pp 92–106
8. Baruque B, Corchado E, Mata A, Corchado JM (2010) A forecasting solution to the oil spill problem based on a hybrid intelligent system. Inf Sci 180(10):2029–2043
9. Bauer B, Huget MP (2003) FIPA modeling: agent class diagrams
10. Bellifime F, Poggi A, Rimasa G (2001) JADE: a FIPA2000 compliant agent development environement. In: Proceedings of the 5th international conference on autonomous agents. ACM, pp 216–217
11. Bergmann R, Wilke W (1996) On the role of abstraction in case-based reasoning. In: Proceedings of the EWCBR-96 European conference on case-based reasoning LNCS, vol 1186, pp 28–43
12. Bratman ME (1987) Intentions, plans and practical reason. Harvard University Press, Cambridge
13. Bratman ME, Israel D, Pollack ME (1988) Plans and resource-bounded practical reasoning. Comput Intell 4:349–355
14. Corchado JM, Lees B (2001) A hybrid case-based model for forecasting. Appl Artif Intell 15(2):105–127
15. Corchado JM, Laza R (2003) Constructing deliberative agents with case- based reasoning technology. Int J Intell Syst 18(12):1227–1241
16. Corchado JM, Aiken J, Corchado E, Lefevre N, Smyth T (2004) Quantifying the ocean's CO_2 budget with a CoHeL-IBR system. In: Proceedings of the 7th European conference on case-based reasoning. LNCS, vol 3155, pp 533–546
17. Corchado JM, Pavón J, Corchado E Castillo LF (2005) Development of CBR-BDI agents: a tourist guide application. In: Proceedings of the 7th European conference on case-based reasoning 2004. LNAI, vol 3155, pp 547–559
18. Chen-Chia C, Zne-Jung L (2011) Hybrid robust support vector machines for regression with outliers. Appl Soft Comput 11(1):64–72
19. Chi-Jie L, Tian-Shyug L, Chih-Chou C (2009) Financial time series forecasting using independent component analysis and support vector regression. Decis Support Syst 47(2):115–125
20. DeLoach S (2001) Anlysis and design using MaSE and AgentTool. In: Proceedings of the 12th midwest artificial intelligence and cognitive science conference (MAICS)
21. Dransfeld S, Tatnall AR, Robinson IS, Mobley CD (2004) A comparison of multi-layer perceptron and multilinear regression algorithms for the inversion of synthetic ocean colour spectra. Int J Remote Sens 25(21):4829–4834
22. Dransfeld S, Tatnall AR, Robinson IS, Mobley CD (2005) Prioritizing ocean colour channels by neural network input reflectance perturbation. Int J Remote Sens 26(5):1043–1048
23. EURESCOM MESSAGE (2001) Methodology for engineering systems of software agents. Technical report P907-TI1, EURESCOM
24. Farquad MAH, Ravi V, Bapi Raju S (2008) Support vector regression based hybrid rule extraction methods for forecasting. Expert Syst Appl 37(8):5577–5589
25. Feret MP, Glasgow JI (1994) Experienced-aided diagnosis for complex devices. In: Proceedings of the 12th national conference an artificial intelligence, pp 29–35
26. FIPA Foundation for Intelligent Physical Agents. http://www.fipa.org. Accessed 20 Mar 2010
27. Fyfe C, Corchado ES (2002) Maximum likelihood hebbian rules. In: Proceedings of the European symposium on artificial neural networks, pp 143–148
28. Glez-Bedia M, Corchado JM, Corchado ES, Fyfe C (2002) Analytical model for constructing deliberative agents. Eng Intell Syst 3:173–185

29. González-Romera E, Jaramillo-Morán MA, Carmona-Fernández D (2008) Monthly electric energy demand forecasting with neural networks and Fourier series. Energy Convers Manag 49(11):3135–3142
30. Herrero Á, Corchado E, Saiz L, Abraham A (2010) DIPKIP: a connectionist knowledge management system to identify knowledge deficits in practical cases. Comput Intell 26(1):26–56
31. http://www.pml.ac.uk/
32. Hussain A, Jaffar MA, Mirza AM (2010) A hybrid image restoration approach: fuzzy logic and directional weighted median based uniform impulse noise removal. Knowl Inf Syst 24(1):77–90
33. Iglesias C, Garijo M, Gonzalez JC, Velasco JR (1998) Analysis and design of multiagent systems using MAS-CommonKADS. Inteligent Agents IV 1365:313–326
34. Javeed SSAK, Al-Garni AZ (1995) Forecasting electric energy consumption using neural networks. Energy Policy 23(12):1097–1104
35. Joh DY (1997) CBR in a changing environment. Case-based reasoning research and development. In: Proceedings of the ICCBR-97, p 126
36. Jeffer CD, Woolf DK, Robinson IS, Donlon CJ (2007) One-dimensional modelling of convective CO_2 exchange in the Tropical Atlantic. Ocean Model 19(3–4):161–182
37. Jeffery CD, Robinson IS, Woolf DK, Donlon CJ (2008) The response to phase-dependent wind stress and cloud fraction of the diurnal cycle of SST and air–sea CO_2 exchange. Ocean Model 23(1–2):33–48
38. Kinny D, Georgeff M (1991) Commitment and effectiveness of situated agents. In: Proceedings of the twelfth international joint conference on artificial intelligence, pp 82–88
39. Kolodner J (1983) Maintaining organization in a dynamic long-term memory. Cognit Sci 7:243–280
40. Kolodner J (1983) Reconstructive memory, a computer model. Cognit Sci 7(4):281–328
41. Kolodner J (1993) Case-based reasoning. Morgan Kaufmann, Los Altos
42. Kolodner J (1993) Maintaining organization in a dynamic long-term memory. Morgan Kaufmann, Los Altos
43. Lavender SJ, Pinkerton MH, Moore GF, Aiken J, Blondeau-Patissier D (2005) Modification to the atmospheric correction of SeaWIFS ocean colour images over turbid waters. Cont Shelf Res 25:539–555
44. Lefevre N, Aiken J, Rutllant J, Daneri G, Lavender S, Smyth T (2002) Observations of pCO_2 in the coastal upwelling off Chile: sapatial and temporal extrapolation using satellite data. J Geophys Res 107(6):8.1–8.15
45. Leake D, Kendall-Morwick J (2008) Towards case-based support for e-science workflow generation by mining provenance. Adv Case Base Reson 5239:269–283
46. Lessmann S, Voß S (2009) A reference model for customer-centric data mining with support vector machines. Eur J Oper Res 199(2):520–530
47. Li H, Liang Y, Xu Q (2009) Support vector machines and its applications in chemistry. Chemom Intell Lab Syst 95(2):188–198
48. Liu S, Duffy AHB, Whitfield RI, Boyle IM (2010) Integration of decision support systems to improve decision support performance. Knowl Inf Syst 22(3):261–286
49. Lo S (2008) Web service quality control based on text mining using support vector machine. Expert Syst Appl 34(1):603–610
50. Lu J, Li R, Zhang Y, Zhao T, Lu Z (2010) Image annotation techniques based on feature selection for class-pairs. Knowl Inf Syst 24(2):325–337
51. Martín FJ, Plaza E, Arcos JL (1999) Knowledge and experience reuse through communications among competent (peer) agents. Int J Softw Eng Knowl Eng 9(3):319–341
52. Melesse AM, Hanley RS (2005) Artificial neural networknext term application for multi-ecosystem carbon flux simulation. Ecol Model 189(3–4):305–314
53. Montaño JJ, Palmer A (2002) Artificial neural networks, opening the black box. Metodología de las Ciencias del Comportamiento 4(1):77–93
54. Nwana HS, Ndumu DT, Lee LC, Collins JC (1999) ZEUS: a toolkit for building distributed multi-agent systems. Appl Artif Intell J 1(13):129–185
55. Nguyen H, Chan W (2004) Multiple neural networks for a long term time series forecast. Neural Comput Appl 13(1):90–98
56. Odell J, Huget MP (2003) FIPA modeling: interaction diagrams
57. Odell J, Levy R, Nodine M (2004) FIPA modeling TC: agent class superstructure metamodel. FIPA meeting and interim work
58. Olivia C, Chang CF, Enguix CF, Ghose AK (1999) Case-based BDI agents: an effective approach for intelligent search on the world wide web. In: Proceedings of the AAAI spring symposium on intelligent agents, pp 22–24
59. OMG Unified Modelling Language Specification. Version 1.3. http://www.omg.org. Accessed 20 Mar 2010

60. O'Reilly JE, Maritorena S, Mitchell BG, Siegel DA, Carder KL, Garver SA, Kahru M, McClain C (1999) Ocean color chlorophyll algorithms for SeaWIFS. J Geophys Res Oceans 103(11):24937–24953
61. Osowski S, Garanty K (2007) Forecasting of the daily meteorological pollution using wavelets and support vector machine. Eng Appl Artif Intell 20(6):745–755
62. Perner P (1998) Different learning strategies in a case-based reasoning system for image interpretation. Advances in case-based reasoning. LNCS, vol 1488, pp 251– 261
63. Perner P (2001) Why case-based reasoning is attractive for image interpretation. Case-based reasoning research and development LNCS, vol 2080, pp 27–43
64. Ping-Feng P, Kuo-Ping L, Chi-Shen L, Ping-Teng C (2010) Time series forecasting by a seasonal support vector regression model. Expert Syst Appl 37(6):4261–4265
65. Pokahr A, Braubach L, Lamersdorf W (2003) Jadex: implementing a BDI-infrastructure for JADE agents. In: EXP—in search of innovation (Special Issue on JADE) 3(3):76–85
66. Rao AS, Georgeff MP (1995) BDI agents: from theory to practice. In: Proceedings of the first international conference on multi-agent systems, pp 312–319
67. Rutgersson A, Smedman A (2010) Enhanced air–sea CO_2 transfer due to water-side convection. J Mar Syst 80(1–20):125–134
68. Sarmiento JL, Dender M (1994) Carbon biogeochemistry and climate change. Photosynth Res 39: 209–234
69. Seung HS, Socci ND, Lee D (1998) The rectified gaussian distribution. In: Proceedings of the advances in neural information processing systems, vol 10, pp 350–356
70. Smola A, Scolköpf B (2003) A tutorial on support vector regression. Stat Comput 14:199–222
71. Takahashi T, Olafsson J, Goddard JG, Chipman DW, Sutherland SC (1993) Seasonal Variation of CO_2 and nutrients in the high-latitude surface oceans: a comparative study. Glob biochem Cycles 7(4):843–878
72. The Beam Project (2010) http://www.brockmann-consult.de/beam/. Accessed 20 Mar 2010
73. Tripathi S, Srinivas VV, Nanjundiah RS (2006) Downscaling of precipitation for climate change scenarios: a support vector machine approach. J Hydrol 330(3–4):621–640
74. Vapnik VN (1999) An overview of statistical learning theory. IEEE Trans Neural Netw 10:988–999
75. Vapnik V (1995) The Nature of statistical learning theory. Springer, Berlin
76. Wendler J, Lenz M (1998) CBR for dynamic situation assessment in an agent-oriented setting. In: Proceedings of the AAAI-98 workshop on CBR integrations, pp 172–176
77. Wooldridge M, Jennings NR (1994) Agent theories, architectures, and languages: a survey. Intelligent agents. LNCS, pp 1–39
78. Wooldridge M, Jennings NR, Kinny D (2000) The gaia methodology for agent-oriented analysis and design. J Auton Agents Multi Agent Syst 3(3):285–312
79. Yang Q, Li X, Shi X (2008) Cellular automata for simulating land use changes based on support vector machines. Comput Geosci 34(6):592–602
80. Zeichen MM, Robinson IS (2004) Detection and monitoring of algal blooms using SeaWIFS imagery. Int J Remote Sens 25(9):1797–1798

Author Biographies

Juan F. De Paz (PhD) Received a PhD in Computer Science from the University of Salamanca (Spain) in 2010. He is Assistant Professor at the University of Salamanca and researcher at the BISITE research group (http://bisite.usal.es). He obtained a Technical Engineering in Systems Computer Sciences degree in 2003, an Engineering in Computer Sciences degree in 2005 at the University of Salamanca and Statistic degree in 2007 in the same University. He has been coauthor of published papers in several journals, workshops and symposiums.

Javier Bajo (PhD) Received a PhD in Computer Science and Artificial Intelligence from the University of Salamanca in 2007. At present, he is Director of the Data Processing Center and Associate Professor at the Pontifical University of Salamanca (Spain) and researcher at the BISITE research group (http://bisite.usal.es) at the University of Salamanca (Spain). He obtained an Information Technology degree at the University of Valladolid (Spain) in 2001 and an Engineering in Computer Sciences degree at the Pontifical University of Salamanca in 2003. He has been a member of the organizing and scientific committee of several international symposiums such as CAEPIA, IDEAL, and HAIS and is coauthor of more than 170 papers published in recognized journals, workshops, and symposiums.

Angélica González (PhD) Received a PhD in Computer Science from the University of Salamanca in 2000. She is currently a Lecturer in Salamanca's University Department of Computer Science and has attended several Master's courses. She is further a professor and tutor for UNED (Universidad Española de Educación a Distancia, Spain's Open University). In the past, she carried out relevant administrative tasks, such as Academic Secretary of the Science Faculty (1996–2000) and Chief of Staff for the University of Salamanca (2000–2003). From 1990, she has cooperated with the Home Ministry, and from 2008 with the Home and Justice Counsel of the local government (Junta de Castilla y León). She is a member of the research group BISITE (http://bisite.usal.es) and has lead several research projects sponsored by both public and private institutions in Spain. She is the coauthor of several works published in magazines, workshops, meetings, and symposia.

Sara Rodríguez (PhD) Received a PhD in Computer Science from the University of Salamanca in 2010. She pursued her studies of PhD in this University. She obtained a Technical Engineering in Systems Computer Sciences degree in 2004, an Engineering in Computer Sciences degree in 2007 at the University of Salamanca. She has participated as a coauthor in papers published in recognized international conferences and symposiums.

Juan M. Corchado (PhD) Received a PhD in Computer Science from the University of Salamanca in 1998 and a PhD in Artificial Intelligence (AI) from the University of Paisley, Glasgow (UK) in 2000. At present, he is Dean at the Faculty of Computer Sciences, Associate Professor, Director of the Intelligent Information System Group (http://bisite.usal.es) and Director of the MSc programs in Computer Science at the University of Salamanca (Spain). Previously, he was sub-director of the Computer Science School at the University of Vigo (Spain, 1999–00) and a Researcher at the University of Paisley (UK, 1995–98). He has been a research collaborator with the Plymouth Marine Laboratory (UK) since 1993. He has led several Artificial Intelligence research projects sponsored by Spanish and European public and private sector institutions and has supervised seven PhD students. He is the coauthor of over 230 books, book chapters, journal papers, technical reports, etc.