Coastal Ecosystems Simulation: A Decision Tree Analysis for Bivalve's Growth Conditions

João Pedro Reis[1], António Pereira[1,2] and Luís Paulo Reis[2,3]
[1]FEUP - University of Porto, Faculty of Engineering - DEI
Rua Dr. Roberto Frias s/n 4200-465 Porto, Portugal
[2]LIACC - Artificial Intelligence and Computer Science Lab., University of Porto
Rua Dr. Roberto Frias s/n 4200-465 Porto, Portugal
[3]EEUM - School of Engineering, University of Minho - DSI
Campus de Azurem 4800-058 Guimaraes, Portugal
Emails: {ei07119, amcp}@fe.up.pt, lpreis@dsi.uminho.pt

KEYWORDS

Decision Trees, Patterns, Data Preparation, Ecological Behavior.

ABSTRACT

The usage of data mining models has the main purpose of discovering new patterns from dataset analysis by extracting knowledge from data and converting it to information. The most challenging part of problem solving is not the generation of high number of instances in dataset, most often hard to understand, but the interpretation of all those instances to extrapolate information about it. Simulation of coastal ecosystems is used to replicate some real conditions related with physical, chemical and biological processes, and produces large datasets from which it could be deduced some information about attributes behaviors. This paper relates the use of Decision Tree models to analyze the growth of bivalve species in an ecosystem simulation. With a set of attributes that represents the water quality in certain modeled regions, the usage of Decision Tree is intended to identify the most significant attribute conditions, which could justify the growth behavior for each analyzed species. This approach aims the creation of new information about how water conditions should be to promote a healthy and fast growth of the analyzed species, being useful to know in which zones the bivalve should be seeded, and which are the conditions that aquaculture producers should afford to benefit the quality of its crops.

INTRODUCTION

Modeling and simulation processes are intended to provide realistic environment for the analysis of a certain simulation. The simulated area that was used to obtain the final results refers to Sango Bay in China, in which were modeled hydrodynamic and biogeochemical variables (Duarte et al., 2003). All these results were provided by the EcoSimNet framework that is a platform for simulation and support decision-making (Pereira et al., 2009).

The large data set analyzed in this study is the simulation result of a lagoon ecosystem, modeled as two-dimensional vertically integrated. Partial differential equations were used to provide dataset attributes that describes different species, water temperature, water quality, seeded cells position, etc. The simulation is based on a finite difference bathymetric staggered grid with 35 lines by 32 columns, generating 1120 cells, with spatial resolution of 500m (Pereira et al., 2009). The aquaculture doesn't use all the 1120 representative cells of the area model - only 352 cells were chosen to seed - and the simulation covers one year and a half of real-time, the bivalve's growth cycle. All this information results in 800 000 instances for the final simulation dataset, being important the implementation of data preparation phases, first to remove not relevant information, and secondly to choose only the pertinent attributes to the analysis.

The dataset used has several attributes, being a set of them a representation of water quality in aquaculture. The subset selection of these attributes is a common problem in the data mining models, due to its improvement of performance (Quinlan, 1996). In the case of Decision Tree models, that are widely used in data mining and decision support applications (Pach and Abonyi, 2006) and specifically the usage of C4.5 algorithm, allows an efficient analysis of continuous variables, which is a characteristic of simulated environments. With this approach, we could obtain a large spectrum of correlated variables that describe how all the water conditions should be to promote a certain growth behavior.

Initially this paper tells about the *State of the art*, in other words, the developed works and studies related with bivalve's growth behavior and its physiology process. Thereafter it will be presented the C4.5 algorithm, which was the algorithm used for the construction of Decision Tree, and the Section *Dataset Preparation* that sets out the preprocessing phases to prepare the dataset used, describing the different phases that composes it. Section *Implementation* refers to the implementation phase that has the purpose to provide the final dataset used to C4.5 algorithm Decision Tree appliance. Section *Experiments* is the section in which experiments are described seeking the best values to apply in its parameters. Section *Results* shows the obtained results, followed by the Section *Conclusions*, in which the results are discussed

Proceedings 26th European Conference on Modelling and Simulation ©ECMS Klaus G. Troitzsch, Michael Möhring, Ulf Lotzmann (Editors)
ISBN: 978-0-9564944-4-3 / ISBN: 978-0-9564944-5-0 (CD)

reaching some conclusions about the approach used and the attributes relations.

STATE OF THE ART

In the last decades many authors studied the ecosystem properties, aiming a better understanding of how bivalves grow and its physiology could be influenced. Authors like Gilbert (Gilbert, 1973), Bachelet (Bachelet, 1980), Appeldoorn (Appeldoorn, 1983) and Beukema (Beukema and Meehan, 1985) claimed, in the early 80's and 90's, that temperature and food quality are two of the most important factors influencing the bivalve physiology. In the late 90's, Smaal & Haas also contributed proving that seston concentrations and chlorophyll-a levels near the bottom are generally higher than surface values, showing the importance of suspended particulate matter - Boundary SPM concentration (Smaal and Haas, 1997). A review made by Saxby in Western Australia, year 2002, that includes several sites like Seto Inline Sea in Japan and Saldanha Bay in South Africa, proves that phytoplankton - high values affects positively - along with nutrients, water temperature and salinity - prolonged exposure to low salinity mays have depressed all growth parameters - affects the bivalve physiology and food quality (Saxby, 2002).

On the other hand, the technology and tools to reach these conclusions are also significant. The author Michel R. Claereboudt used the GMDH algorithm (Group Method of Data Handling) to achieve some of these results. This algorithm is an inductive process that selects the best solutions for a given problem using the *external criterion* (Claereboudt, 1994). Despite the distinct approach of this algorithm and C4.5 (Quinlan, 1992), both reached good results in this context.

C4.5 ALGORITHM

The most common utilization of Decision Trees (DT) lies in the classification of instances from well-known datasets. There are two types of datasets that are mandatory in classification method, which are the training dataset, and the test dataset. The first one lies about the instances that are already classified, and the second tells about the instances that have to be classified based on the training dataset. For the construction of DT, all the attributes are tested as root nodes using a criterion for split, that can be a binary or a multi-way split, until the most informative attribute is found. The dataset is then divided by the root node split, and recursively subsequent trees are calculated, utilizing the partitioned dataset, until all samples for a given dataset belong to the class.

Decision Trees are attractive in the utilization of data mining models due to its intuitive representation (easy to understand by humans), its relatively fast construction, compared with other models, and its comparable or superior accuracy to other models. The utilization of a Decision Tree implies an algorithm that guides the construction of nodes and leafs. The nodes of DT depicts the attributes of dataset, and leafs represent the labeled instances, being the C4.5 algorithm chosen for its construction (Quinlan, 1992). This algorithm is one of best-known and most widely-used in learning models, allowing the analysis of numeric attributes, which is the case of our problem. The split criterion used in this problem is the Gain Ratio, instead of Information Gain, due to the high levels of entropy (common in continuous variables). The choice of this criterion is well explained in the *Implementation* Section.

In the problem solving method developed, the main purpose was to take advantage of Gain Ratio criterion, knowing the split nodes (conditions of the attributes analyzed) and values of attributes that reach the classified instances. Hence, was only used a set of classified instances to produce a DT based on it, with the intent of capture the Water Conditions of each labeled instance.

DATASET PREPARATION

The analysis of the dataset to be used, is an important phase that could be done by some essential steps. CRISP-DM (Shearer, 2000) is a Data Mining Process that aims in dataset analysis, on a specific domain of problem. For the growth behavior analysis, only the *Data Preparation* phase of this model was used, which is constituted by sub phases like *Data Selection, Data Cleaning, Data Construction, Data Integration and Data Formatting*. Only this phase was used due to the fact of merely clean and construct data phases were needed, making the processed dataset capable of being analyzed, and for further implementation of C4.5 algorithm.

In the first phase of the process, *Data Cleaning*, the treatment is focused on outliers - data that is not common or expected to be different, that in this case represents a non-seeded cell - and missing data. The treatment of these cases is positively important to the final result, in which a consistent analysis couldn't be made due to the corrupted data and miss representation of information. The used dataset from EcoSimNet framework, like previously said, is composed by a 32 lines by 35 columns, but only 352 cells represent the total number of seeded species, due to the existence of land cells and boundaries of the ecosystem. These specific cases of non-seeded cells could be easily found, since a very high number of variables, e.g. shell length, was used. The option was to remove the outlier's cells, being this information not relevant to the final analysis. To the amount of instances produced by the simulation of 731 iterations (days), the removal of missing or corrupted data seems not very harmful. The amount of instances per cells remains sufficient to make a posterior good analysis.

The Subset Selection Problem is a very common problem of attribute selection for dataset analysis. This attribute selection is totally relevant for our solution, since the analysis is focused on the Water Quality, and not in the whole information of the dataset. As we are dealing with continuous variables - Modeling and Simulation

(John et al., 1994) - the appliance of C4.5 algorithm is adequate, and the attribute selection promotes its performance.

The concept of Entropy may be informally defined as the measure of impurity in a group example. It is maximum when we cannot predict nothing from the data - the probability of choosing an example in a group is the same - and it is minimum when we can say for sure that a certain data will be chosen - the probability of choosing an example is 1 (only one type of data in the group). This concept is important, because the several data regarding the dataset have a high level of entropy that is a characteristic of continuous variables. Due to the *Data Cleaning* phase, the value of entropy was significantly reduced, improving the efficiency of the work, being the final result more consistent and credible.

After this step, we have to be aware of the attributes that are important to achieve the main final purpose. The initial dataset has the following attributes, excluding the time step, position, and species information: *Box depth*: depth of the seeded box; *Dynamic height*: height of the water in a determined cell as tide's result; *U Velocity*: velocity of the water in the longitude orientation; *V Velocity*: velocity of the water in the latitude orientation; *Salinity*; *DIN*: Dissolved Inorganic Nitrogen; *Phytoplankton biomass*; *POM*: Particulate Organic Matter; *TPM*: Total Particulate Matter; *Water temperature*; *Zooplankton biomass*; *Boundary NO3 concentration*: nitrate - indicator of water quality; *Boundary POM concentration*: Particulate Organic Matter; *Boundary SPM concentration*: Suspended Particulate Matter; *Boundary Zoo concentration*.

After an analysis phase, in which we select the attributes that are significant to the problem, the final selection attributes are the following: *Boundary NO3 concentration*; *POM*; *Phytoplankton biomass*; *Boundary SPM concentration*. All these attributes represent the quantity of particulate matter and the level of pollution in the water in which the bivalves are seeded.

The last step of this phase, is to separate the species information for a further independent treatment. In the EcoSimNet simulations, it was used three types of species: Chlamys Farreri (scallops), Crassostrea Gigas (oysters) and Laminaria Japonica (algae).

When the *Data Preparation* phase is concluded, the main question that have to be made, regarding the main goal of the problem, is: *how can water quality influence the bivalve's growth?* Firstly we have to analyse the growth behavior of some cells, in order to consider if an improvement of growth could be made. Figure 1 is a representation of Chlamys Farreri growth behavior (Scallop), from seed to harvest season, being each line a single cell of the seeded grid simulated.

As can be seen from Figure 1, not all the cells have the same behavior pattern, or even the same final shell length when the harvest season occurs. This is a great indicator to deduce the water conditions that promotes a good bivalve growth. One of the purposes of this work is to answer the question above, with information that can be understood by a regular person, provided by very intuitive and easy seen representation methods.

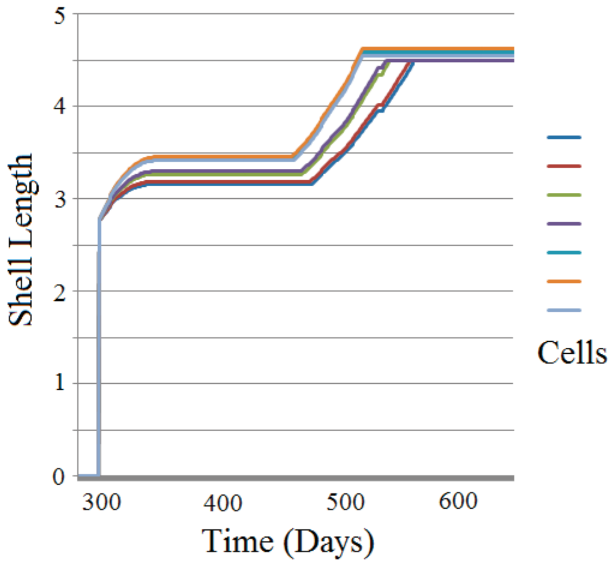

Figure 1: Chlamys Farreri Growth Behavior

Laminaria Japonica won't be contemplated for further analysis due to the lack of unusual and specific conditions that could be explored to induce a good growth. Hence, only Chlamys Farreri and Crassostrea Gigas will be used to generate its correspondent Decision Tree, with the C4.5 algorithm usage.

Using the derivative of growth species, we could classify the instances of simulation into Good and Bad growth. This *Quality Measure* sub-section is intended to establish a value (*Threshold*) that should separate these two classifiers, in order of being capable to distinguish the attribute circumstances that promote a certain growth, in the final phase of this work. As previously said, the growth derivative was used, represented by Figure 2, being these values a representation of growth registered in a certain time step of the simulation - slope between two neighbor time steps.

Which threshold represents better a quality measure? It is the question that should be answered to reach a high confidence in each species dataset analysis, in which the number of instances per classification Good or Bad has to be considerable and balanced. So, if the number of Good classified instances is significantly low compared with Bad classified ones, we could say that the dataset could be biased due to the unbalanced number of instances, and benefit one of the classifications. Hence, the *Threshold* - value that separates the derivative function of each growth species, into a good or bad label - will be discussed in the following sections.

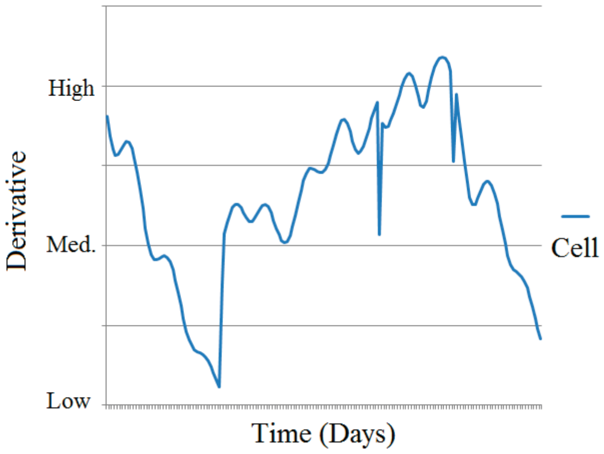

Figure 2: Chlamys Farreri Derivative Growth Behavior

IMPLEMENTATION

The utilization of C4.5 algorithm lies in the fact of dataset attributes generated from the EcoSimNet simulation process being continuous. This algorithm was improved and developed with different versions along time, and now is capable of treat continuous attributes efficiently (Quinlan, 1996).

Relatively to the C4.5 algorithm parameters, the criterion used was the Ratio Gain due to its more adequate appliance, comparing to Information Gain. The fact of dealing with high different values per attribute, normally found in continuous variables (Witten and Frank, 2005), like numeric values, the *Information Gain* approach could be biased, and overfitting could occur - selection of non-optimal attribute for prediction. The *Gain Ratio* is a based *Information Gain* method that takes into account the number of attribute instances, reducing the bias on high-branch attributes. The *Information Gain* is a based *Entropy* method that takes into account the lower value of entropy, high information gain value, to choose the root of the calculated tree.

This type of criterion is used to calculate the root of a tree that maximizes the *Ratio Gain*. Hence, the final tree is the result of an iterative process that calculates the next attribute to use, taking into account the previous one. So, while the tree is constructed, the number of instances is reduced due to the fact of previous attributes limitations and produced leafs. Hence, with this parameters it is possible to modulate a consistent *Decision Tree* that fulfills the goals of this project.

EXPERIMENTS

The *RapidMiner 5* was the framework used to run the experiments with different C4.5 algorithm parameters. It is the most Data Mining framework tool used (KDnuggets, 2010), allowing the creation of Data Mining models, and using the implementations of most relevant algorithms for different types of domains, like classification, clustering and item set mining.

The *Minimal Gain* parameter is the minimum value of Gain Ratio that should occur in an attribute to be chosen for tree expansion, and *Minimal Leaf Size* parameter is the minimum number of instances that a leaf should have in the Decision Tree. These were the two parameters tested to reach a good Decision Tree representation for growth conditions deduction. These parameters variation produces different numbers of nodes and leafs in Decision Trees, and should neither be too high, nor too low, due to the difficult interpretation of its representation.

Chlamys Farreri and Crassostrea Gigas species were tested, and the relation between the Number of Nodes, Minimal Gain and the Minimal Leaf Size will be presented.

Related to Chlamys Farreri species, it can be seen from Figure 3 that best results are provided from the variation of Minimal Leaf Size value between 25 and 100 and Minimal Gain value equal to 0.01 or 0.03. Value of Minimal Leaf Size equal to 10 with Minimal Gain equal to 0.03 and 0.01 produces a very high number of nodes, which is neither a good visual representation, neither easy to deduce the conditions that promote a certain growth.

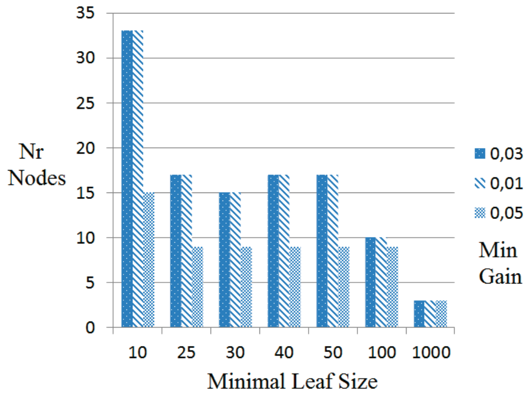

Figure 3: Chlamys Farreri: Number of Nodes

From figure 4 it is easy to see that the best results of Crassostrea Gigas Decision Tree are for Minimal leaf size between 30 and 50, with any value of Minimal Gain. A good result could also be obtained with Minimal Leaf Size value equal to 10, but only with Minimal Gain equal to 0.001.

RESULTS

Each species has its own growth information in separated datasets. For each species analysis, it will be presented the relations of dataset labeled attributes.

An important observation that has to be made before the tables analysis (Tables 1 thru 4), is that low values of instances per conditions (set of rules that satisfy a certain label: Good or Bad), do not discard the confidence inherent to it. The purpose of this paper is to find the circum-

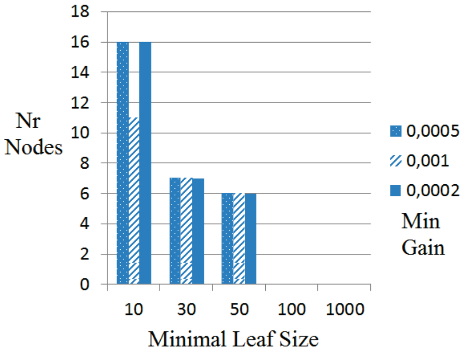

Figure 4: Crassostrea Gigas: Number of Nodes

stances that could determine a certain growth, independently of their number of instances. We are looking for the specificity that water conditions could provide, separating the two labels of associated growth. Another consideration that should be made, is relatively to maximum and minimum values, generated by the simulation framework, from each attribute analyzed. The presented tables will not represent these values, so we assume that when a situation like *NO3 < 4.5* and *NO3 > 4.5*, the NO3 values should variate between its minimum possible value and 4.5, and between 4.5 and its maximum possible value, respectively.

This analysis had two different intentions: Make a qualitative evaluation of most significant attributes, and compare two totally independent analysis: the Decision Tree results, in which could be validated the computer results, with the dataset visual analysis. Due to length restrictions of the paper only the computer analysis will be done.

Chlamys Farreri

As said earlier, a dataset was generated to obtain this type of results. The number of instances that dataset contains is 15 042, being 7724 (51%) labeled as Bad Growth, and 7318 (49%) labeled as Good Growth. This dataset follows the Java implementation metrics, being: Threshold=0.015 and the Derivative Factor=1.

Figure 5 represents the Chlamys Farreri Decision Tree, with the parameters: Minimal Size for Split=2, Minimal Leaf Size=40, Minimal Gain=0.01, Maximal Depth=20 and Confidence=0.25.

From Figure 5 an analysis was made having originated two different tables. One of them tells about a Good growth conditions, Table 1, and the other Bad growth conditions, Table 2.

Regarding Table 1, if Nitrate (NO3) levels are above the 4.586 and Phytoplankton above 0.138 a good growth will occur. To ensure this growth, the POM values should be above 2.218. These water conditions should benefit the Chlamys Farreri species growth.

Regarding Table 2, if NO3 levels are between 1.093 and 4.586, a bad growth will occur. To ensure this growth, the POM value should be below 2.218. These water conditions should worsen the growth of Chlamys Farreri species.

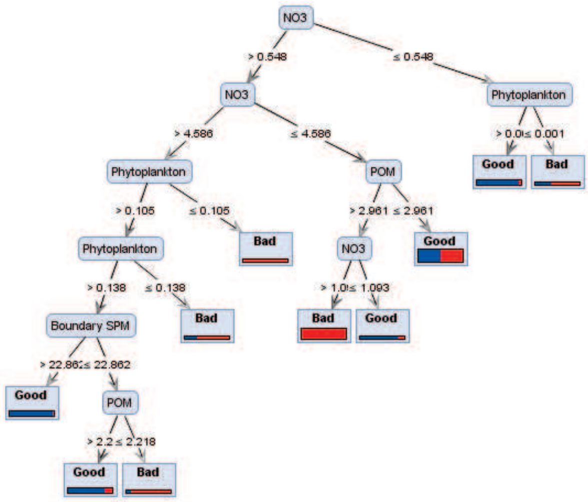

Figure 5: Chlamys Farreri Decision Tree

Table 1: Good Growth: Chlamys Farreri

Leaf	Confidence	Total	Conditions
Good	98.0	1419	$NO3 > 4.586$ $Phytoplankton > 0.138$ $BoundarySPM > 22.862$
Good	94.6	1499	$NO3 \leq 0.548$ $Phytoplankton > 0.001$
Good	84.0	1514	$NO3 > 4.586$ $Phytoplankton > 0.138$ $BoundarySPM < 22.862$ $POM > 2.218$

Crassostrea Gigas

The next analysis lies in the Crassostrea Gigas species, and dataset description. The number of instances of the dataset is 19 991, being 10 397 (52%) labeled as Bad Growth, and 9594 (48%) labeled as Good Growth. This dataset follows the Java implementation metrics, being: Threshold=0.02 and the Derivative Factor=1.

Figure 6 represents the Crassostrea Gigas Decision

Table 2: Bad Growth: Chlamys Farreri

Leaf	Confidence	Total	Conditions
Bad	99.3	467	$NO3 > 4.586$ $Phytoplankton = 0.138$
Bad	98.6	4006	$1.093 < NO3 < 4.586$ $POM > 2.961$
Bad	87.3	133	$NO3 > 4.586$ $Phytoplankton > 0.138$ $BoundarySPM \leq 22.862$ $POM \leq 2.21$

Table 3: Good Growth: Crassostrea Gigas

Leaf	Conf.	Total	Conditions
Good	83.3	48	$POM = 0.949$ $NO3 \leq 0.336$
Good	64.5	31	$POM = 0.949$ $NO3 > 0.336$ $Phytoplankton > 0.297$ $BoundarySPM > 16.171$
Good	54.6	18787	$0.949 < POM \leq 4.630$

Table 4: Bad Growth: Crassostrea Gigas

Leaf	Conf.	Total	Conditions
Bad	100.0	69	$POM \leq 0.725$
Bad	99.2	258	$POM = 0.949$ $NO3 > 0.336$ $Phytoplankton \leq 0.297$
Bad	88.4	190	$POM > 4.630$

Tree with the parameters: Minimal Size for Split=2, Minimal Leaf Size=10, Minimal Gain=0.001, Maximal Depth=20 and Confidence=0.25.

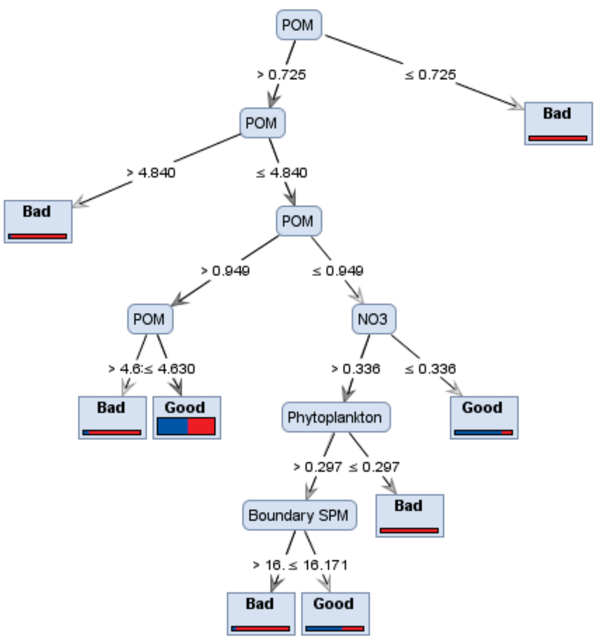

Figure 6: Crassostrea Gigas Decision Tree

Two different tables were generated when analysing Figure 6. One of them tells about a Good growth conditions, Table 3, and the other Bad growth conditions, Table 4.

Regarding Table 3, if POM values are between 0.949 and 4.630 a good growth will occur. To ensure this growth, the SPM values should be below 16.171. These water conditions should benefit the Crassostrea Gigas species growth.

Regarding Table 4, if POM values are higher than 4.630 and lower than 0.725, a bad growth will occur. Low values of phytoplankton reinforce this growth. The reported water conditions will worsen the growth of Crassostrea Gigas species.

CONCLUSIONS

As bivalves are marine and freshwater mollusks, it is obvious that with more pure water - less solid particulates - the bivalves have better conditions to growth healthy and rapidly. By the analysis made previously, it can be deduced that water conditions influence the growth of bivalves.

This approach aims the creation of new information about how water conditions should be to promote a healthy and fast growth of the analyzed species, being useful to know in which zones the bivalve should be seeded, and which are the conditions that aquaculture producers should afford to benefit the quality of its crops.

NO3 levels and POM values are important attributes for the Chlamys Farreri species growth. These two indicators should be controlled to provide guidance to bivalve physiology. High levels of water quality (NO3) and low levels of particulates (water pollution) are needed to produce a beneficial growth. We can see that relatively to the growth of this species, the levels of NO3 are inversely related.

POM and Phytoplankton are the most important attributes for the Crassostrea Gigas species growth. We can conclude that some medium values of POM can benefit its physiology, as much as high values of Phytoplankton. This medium values of POM could be justified by some oyster species life cycle are composed by an attachment to a rocky surfaces, that have some levels of impurity. Phytoplankton describes the environment quality that are important to any species growth, as Crassostrea Gigas.

The results of this study don't accrue only for the bivalve production, but also for the public health. Since the bivalve make part of humans feeding habits, the more quality these species have, the more quality life we are giving to people. With this work, producers can evaluate better the water conditions that promote the species physiology, together with water quality for the ecosystem health and the final consumer.

The application of Decision Tree shown to be a powerful tool to attribute analysis and its behavior and relation with other attributes, and more specifically, the C4.5 algorithm. Also, the Decision Tree is a very intuitive way to deduce the attribute behavior, and representation tool due to its simple and direct presentation. The C4.5 algorithm has demonstrated to be a powerful tool in datasets of continuous attribute, having different flexible parameters that can provide a better solution comparing with different approaches from Decision Tree based.

REFERENCES

Appeldoorn, R. S. (1983), *Variation in the growth rate of Mya arenaria and its relationship to the environment as analyzed through principal components analysis and the omega parameter of the von Bertalanffy equation*. Fishery Bulletin 81: 75-84.

Bachelet, G. (1980), *Growth and recruitment of the tellinid bivalve Macoma balthica at the Southern limit of its geographical distribution, the Gironde estuary (SW France)*. Marine Biology 59: 105-117.

Beukema, J. J. and Meehan, B. W. (1985), *Latitudinal variation in linear growth and other shell characteristics of Macoma balthica*. Marine Biology 90: 27-33.

Chen, Q. and Mynett, A. (2009), *Rule-Based Ecological Model*. In: Handbook of Ecological Modelling and Informatics: 307-324.

Claereboudt, M. R. (1994), *GMDH algorithm as a tool for bivalve growth analysis and prediction*. ICES Journal of Marine Science 51: 439-445.

Debeljak, M. and Deroski, S. (2011), *Decision Trees in Ecological Modelling*. In: Modelling Complex Ecological Dynamics: an Introduction into Ecological Modelling, vol. 2: 197-209.

Duarte, P., Meneses, R., Hawkins, A. J. S., Zhu, M., Fang, J., and Grant, J. (2003), *Mathematical modelling to assess the carrying capacity for multi-species culture within coastal waters*. Ecological Modelling, 168 (1-2): 109-143.

Gilbert, M. A. (1973), *Growth rate, longevity and maximum size of Macoma balthica (L.)*. The Biological Bulletin 145: 119-126.

Hahsler, M., Chelluboina, S., Hornik, K. and Buchta, C. (2011), *The arules R-Package Ecosystem: Analyzing Interesting Patterns from Large Transaction Data Sets*. Journal of Machine Learning Research 12(Jun): 2021-2025.

John, G. H., Kohavi, R. and Pfleger, K. (1994), *Irrelevant Features and the Subset Selection Problem*, MACHINE LEARNING: PROCEEDINGS OF THE ELEVENTH INTERNATIONAL: 121-129.

Jorgensen, S. E. and Bendoricchio, G. (2001), *Fundamentals of Ecological Modelling*.

KDnuggets (2010), *Data Mining/Analytic Tools Used Poll (May 2010)*. Available online: http://www.kdnuggets.com/polls/2010/data-mining-analytics-tools.html.

Pach, F. P. and Abonyi, J. (2006), *Association Rule and Decision Tree based Methods for Fuzzy Rule Base Generation*. Word Academy of Sciences, Engineering and Technology 13: 45-50.

Pereira, A., Reis, L. P. and Duarte, P. (2009) *EcoSimNet: A Multi-Agent System for Ecological Simulation and Optimization*. EPIA 2009, LNAI 5816: 473-484, Springer-Verlag.

Quinlan, J. R. (1992), *C4.5: Programs for Machine Learning*. Morgan Kaufmann.

Quinlan, J. R. (1996), *Improved Use of Continuous Attributes in C4.5*, Journal of Artificial Intelligence Research 4: 77-90.

Saxby, S. A. (2002), *A review of food availability, sea water characteristics and bivalve growth performance at coastal culture sites in temperate and warm temperate regions of the world*. Fisheries Research Report No. 132, Department of Fisheries, Western Australia.

Shearer, C. (2000), *The CRISP-DM Model: The New Blueprint for Data Mining*, Journal of Data Warehousing vol.5, N.4: 13-22.

Smaal, A C and Haas, H A (1997), *Seston Dynamics and Food Availability on Mussel and Cockle Beds*. Estuarine, Coastal and Shelf Science 45: 247-259.

Witten, I. H. and Frank, E. (2005), *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Series in Data Management Systems, Elsevier.

AUTHOR BIOGRAPHIES

JOÃO PEDRO C. REIS was born in Porto, Portugal, and obtained his degree in Informatics Engineering at the University of Porto. Since 2010 he is dedicated to research in the field of Machine Learning and Multi-agent Systems, and he is a MSc student in Informatics and Computation in the University of Porto. His e-mail address is: `ei07119@fe.up.pt`; `joaoreis.correia@gmail.com` and his personal webpage at `http://www.fe.up.pt/~ei07119`.

ANTÓNIO PEREIRA was born in Porto, Portugal, and has a PhD in Informatics Engineering from University of Porto. Since 2003 he is dedicated to research in the field of Agent-Based Simulation, Distributed Artificial Intelligence, Optimization and Intelligent Systems. His email is `amcp@fe.up.pt` and his personal webpage at `http://www.fe.up.pt/~amcp`.

LUÍS PAULO REIS was born in Porto, Portugal, and has a PhD in Electrical Engineering (Coordination in Multi-Agent Systems) in the University of Porto. He has been researching in the area of (Multi-Agent) intelligent simulation for several years in different projects including FC Portugal simulated robotic soccer team World and European champion of RoboCup in 2000 and 2006. His email is `lpreis@dsi.uminho.pt` and his personal webpage stays at `http://www.fe.up.pt/~lpreis`.