A system Architecture for Collaborative Environmental Modeling Research

Philip Sallis\textsuperscript{a}, Subana Shanmuganathan\textsuperscript{a}, Leopoldo Pavesi\textsuperscript{b} and Mary Carmen Jarur Muñoz\textsuperscript{b}

\textsuperscript{a}Auckland University of Technology, New Zealand
\textsuperscript{b}Universidad Catolica del Maule, Chile
philip.sallis@aut.ac.nz

ABSTRACT

This relates to early stage research that aims to build an integrated toolbox of instruments that can be used for environmental modeling tasks. The application area described is grape growing and wine production. A comparative study including data gathered in both New Zealand and Chile is described. Using both passive and sensor technology data is gathered from atmosphere, vines, and soil. Human sensory perceptions relating to wine taste and quality is also gathered. The project proposes a synthesizer which collects and analyzes data in real time. Computational neural network modeling methods and geographic information systems are used for result depiction. This convergence of computational techniques and information processing methods is proposed as being an example of software and systems collaboration. The project called Eno-Humanas is so named because of the blend of the precise enological data and less qualitative human perception data. It is expected that the discrete input elements of the architecture here will be demonstrably dependency-related and derived from correlation values once data gathering instruments and analytical software have been developed. At this stage of the project, these tools and methods are being built and tested. This is the first stage of the project and the proposed research that will come from it in order to answer wide questions such as the ordinal set of data values necessarily present to predict climate conditions, the relationship between vine sap rise and dew point calibrations, towards addressing the popular question of ‘what makes for a good year for wine’. In addition to the bringing together of various technologies, methods and kinds of data, (geo-referential, climatic, atmospheric, terrain, plant biological and qualitative sensory expressions), the paper also describes an international research collaboration and its parameters.

Key words: Hardware device and Software.

1. INTRODUCTION

It is far from unique to employ several different elements of software, hardware and mixed method analysis to solve a problem requiring computation. Nonetheless, the elements required for a particular application domain requires a work-flow architecture that combines information processing and computational methods that does produce a context-specific design. In the case of the project described in this paper, we have conceived a combinatorial approach that utilizes several technologies and methods in order to build a set of tools that when taken together can provide a computational platform for environmental modeling applications. The study to which the project refers is based on an international collaboration between computer scientists and engineers in New Zealand and Chile, bringing together other disciplines and experts from enology, viticulture, climatology, atmospheric science and plant biology. This is assumed to be a long-term project involving many individuals from disparate disciplines and centering on a large data set of different kinds of data; some numeric and precise ranging to text based and qualitative. Motivated by some first-step research that indicates a large investigation base relating to climate and soil condition for example, as determinates of wine quality, this project conceived the creation of a comprehensive data set that can be mined and dependencies derived and extracted from correlations between instance values in the set. We have put this project into the context of Environmental Modeling with the objective of demonstrating the influence that computational methods can have upon that area, particularly in this example application domain.

Environmental modeling takes different forms but is generally regarded as the development of formal depictions, often using mathematical expressions, of variables and parameters that describe the interaction between and the inter-dependence of component
influences on a state of life; be it human, living organism or plant. The research described here relates to environmental influences on grape growing, which as a topic has occupied a substantial segment of the literature pertaining to published scientific work in environmental impact studies, environmental and viticulture planning and management and indeed, in the evolving integrating disciplines of Geo-informatics, Bio-informatics and Bio-economics [1].

The research methodology employed in this research can best be described in terms of a framework for collaborative investigation, overlaid by a systems architecture that is essentially one of component integration, utilizing techniques, software and hardware from numerous sciences. The principal disciplines referenced are Computer Science, Geodesy, Plant Physiology, Geology and Atmospheric Science but a great deal of environmental modeling research has also been undertaken by those working with Geographic Information Systems (GIS) [2].

2. THE RESEARCH BACKGROUND

The interest for the investigators in this research comes from four main directions:

First, a growing relationship between researchers in two universities, one in New Zealand and one in Chile. These two countries share an economic free trade agreement and indeed, a long history of collaboration, particularly in the fields of pastoral farming, agriculture, horticulture and forestry. In recent years numerous joint projects have merged in the area of grape growing and wine production...viticulture and enology. Scientists in both countries and in the universities where the authors of this paper reside have been sharing knowledge and wisdom across a spectrum of research questions that arise from the endeavors of both countries to work together. As it happens, the authors have designed a research project that is both collaborative in terms of effort contribution and also comparative in the sense of data values that are being collected and analyzed for the science being undertaken.

Second, is the identification of a set of parameters and variables that make a comparison between these two countries interesting from a data correlation perspective. The two locations being studied are both on the same latitude: Auckland New Zealand is at 36'S and Talca Chile is at 35'S. Both are in the Southern Hemisphere and therefore, experience the same season each year at the same time. Both countries are acutely aware of the atmospheric ozone layer, which is thin and is receding in the sky above them. Both use similar grape growing and wine production techniques. The land mass is different, with the continental influence of South America impacting on Chile’s climate and especially in the Central Valley topography of the Maule Region where the city and district of Talca is located, while the maritime climate of New Zealand and especially the Auckland Region is influenced by its Island geography and massive surrounding oceans. So potential correlations based on atmosphere, climate, terrain, soil, grape variety and wine production techniques, encouraged the authors to evolve the concept of this research project.

A large number of variables have been used in previous studies to identify specific issues relating to crop production. In focusing on the effect of soil condition for optimal grape growing for example, a New Zealand study produced a table illustrating good and bad sub soil composition, which reports a range of values in their sample where the good sub soil is a meaningful predictor of better potential growth results [3].

Third, although considerable data already exists relating to climatic influences of grape growing for instance, the researchers have proposed a combinatorial approach that is inclusive of a wider spectrum of variables than in the past. This data set includes precise data values collected from atmospheric, climatic, terrain and plant sources but also includes imprecise data values gathered from human sensory perceptions of taste, flavor, complexity and structure of the wine that is produced. This is usually referred to as fuzzy data and comprises vector elements capable of being described in mathematics as rough sets [4]. The investigators are attempting to build scenarios that can be useful is asking the long standing question, “What makes for a good year for wine?”

Fourth, by combining real time precise data collected by atmospheric and other sensors with imprecise data gathered from human sensory perception, it is proposed to build a computational synthesizer to blend these values into a neural network that can adapt its output scenarios according to new information as it is received and processed. Standard parametric statistical techniques are used for hypothesis testing and population analysis but some non-parametric methods are increasingly becoming used with contemporary computational techniques using neural networks for scenario building and even prediction [5]. We are adopting this broader spectrum of analytical techniques for our research methodology. Computational Neural Networks have a growing relevance to and appropriateness for many instances of this kind of combinatorial data mix [6].
It is this fourth element of the research that is considered novel by the authors because it combines historical data and methods of analysis with new real time data collection and processing methods, which blends precise and imprecise data in an endeavor to contribute to the primary quest for an understanding of what makes for a good year for wine.

3. SYSTEM ARCHITECTURE

The researchers have established climate recording instruments in two locations; one in New Zealand and one in Chile. These instruments replicate the data collection of standard meteorological equipment but operate in real time and using wireless telemetry they both submit the collected data to a common server, which publishes the input to the web in one minute intervals (see www.geo-informatics.org). An early analysis of this data for climate trends has been undertaken and the results showing comparisons of the two sites are also being published to the web. Additionally, a prototype weather prediction algorithm has been designed and implemented using a neural network simulator. These results are also published to the web in hourly intervals. This algorithm is described in Section 6.

Sensors used to gather atmospheric influences such as carbon emission saturation levels and herbicide and other air born pollutants are currently being developed for integration with the system architecture. Soil and plant sensors have already been developed and a description of these appears in section 5.

Additional to the real time data being captured, the system architecture ingests historical data from various sources. One such source is the CRIA (Centro de Información Agrometeorológica), by "Ministerio de Agricultura" (Agricultural Minister) [7]. They provide climate data relating to the Maule Region for the years 1995 to 2007. This and other similarly historical data is used by the neural network synthesizer to ‘learn’ about the grape growing influences for scenario building and prediction.

Finally, experiments have yet to be designed to capture the human sensory perception data but the parameters conventionally identified for the assessment of quality in wine and the variables relating to it are known to be structure, complexity, taste and flavor…with refinements of each being within bands of such opinion as granularity in dry or sweet taste and so on. The researchers are now building a fuzzy classification scheme in order to label every variant of opinion in such a form that they can be ingested by the neural network synthesizer for inclusion in the scenario and prediction matrix.

4. AN ILLUSTRATION OF THE SYSTEM ARCHITECTURE

The diagram in Figure 1 depicts the system architecture for this research, illustrating data flows and component interactions. This diagram makes explicit the collaborative dimension between discrete technologies with their underlying science, which when taken together with the international collaborative nature of the project, suggests a multi-faceted approach to environmental modeling that could be generalized to other contexts but which is specific in this example to the grape growing and wine production industries.

5. SENSOR TECHNOLOGY

The sensor technology and telemetry configurations used for this research emanate from previous work by
Previous work by some of the authors with implanted sensors has focused on the rate of sap rise in grape vines (figure 3). A number of articles describe some methodologies to measure sap flow. Different physical and thermodynamic principles are used to determine low and reverse rates of sap flows in woody plants, whereas both invasive and noninvasive procedures have been proposed. In fact, several manufacturers have developed a variety of devices to conduct real time measurements handling very large amounts of data. Examples of noninvasive systems are based on optical, magnetic resonance imaging (MRI), ultrasonic Doppler and laser pulse technologies. Invasive options are on the other hand generally based on thermal balance, which uses heat as a tracer, the heat pulse methods (HPM) being the most known and used in this category [8 9 10]. HPM methods require a three terminal probe so that an invasive measurement based on a stimulus/response process is carried out. The described procedure implies an “intelligent” sequence to be performed, which is normally accomplished by a piece of general purpose data logger equipment. HPM has been used for the last 50 years and makes use of the basic Marshall equation (1958):

\[
\rho c_v \frac{dT}{dt} = \frac{d}{dx} \lambda_x \frac{dT}{dx} + \frac{d}{dy} \lambda_y \frac{dT}{dy} - a \rho c_v \frac{dT}{dx} + Q
\]

which delivers the temperature rise \( T \) at a distance of \( r \) according to

\[
T = \frac{Q}{4\pi \kappa I} e^{-\frac{(x-x_0)^2 + y^2}{4\kappa t}} , \quad r = \sqrt{x^2 + y^2}
\]

The heat pulse velocity can be obtained from

\[
V_Z = \frac{x_p + x_L}{2t_Z}
\]

In 1981, Cohen proposed an improved alternative to HPM that relies on the time needed to achieve a maximum temperature recorded by a single sensor, which is called the T-max method, and is given by

\[
V_M = \frac{\sqrt{x_p^2 - 4\kappa t_M} - 4\kappa t_M}{t_M}
\]

Besides the problem of choosing the most appropriate sensor and measuring method, our research work also focuses on the required interconnection of a network having a large amount of single sensor units, which may as a whole deliver data that can be conveniently used for establishing a model to evaluate the global situation in the area under study. A smart sensor option incorporating local processing power such as a last generation microcontroller is a suitable alternative for this purpose [11]. By using this approach, the clusters of geo-referenced measuring points can be centrally organized and managed from a geographical information system, allowing a context sensitive analysis of real data in a collaborative environment.
An illustration of this software and its use in this investigation can be observed in the weather prediction example that appears on the project’s website. Briefly described it includes elements of time series analysis and other conventional statistics combined with techniques that are collectively referred to a Connectionist Science or Connectionism, to compare this area of Artificial Intelligence with the area known as Symbolic AI. Figure 5 below shows output from the neural network analyzer “learning” from real time climate data to predict upcoming weather conditions. This is a prototype and further variables will be added in due course. The computed results are examined daily and through intervention by the researchers, the values are adjusted for the actual conditions. Hence, the neural network learns over time and the accuracy of its predictions improves.

In the Neucom software multi-layer perceptrons (MLP) are used to generate approximations for end results based on iterations of the analyzer through the data provided in real time from our meteorological instruments in both New Zealand and Chile. As the specification for Neucom states, Multi-layer perceptrons (MLPs) are standard neural network models for learning from data a non-linear function that discriminates (or approximates) data according to output labels (values). MLPs are trained with the use of the back-propagation algorithm developed by Rumelhart et al., 1986. This module of Neucom (KEDRI) creates a 2-layer feed-forward MLP network. It uses linear output unit activation function and scaled conjugate gradient optimization or quasi-newton optimization. The user can specify the number of hidden units/nodes and the number of training cycles and its activation function [12].

The MLP is composed by an input layer (x), a hidden layer (y) and an output layer (z). Each input’s node of the net is connected to hidden layer by a weight, w, and its bias is called δ. Each hidden node is connected to the output layer by a weight w, and its bias is called δ. Each node of the output is calculated by

\[ z_i = \sum w_{ij} y_j - \delta_i = \sum w_{ij} f \left( \sum w_{ij} x_i - \delta_j \right) - \delta_j \]

where f(·) is the activation function each one could be linear, sigmoid, log-sigmoid, etc.

---

In this first example we analysed relative pressure, outdoor temperature and outdoor humidity data collected from the La Crosse WA2300 device installed at the Universidad Catolica del Maule (UCM) (see figure 4). The data labels in the SOM (figure 6) depict the clustering patterns/trends within the daily climate data being analysed.

In our second example discussed herein from [14:p2] we observed the potential for using Kohonen’s SOM method for analyzing wine flavors and attributes related to their geo-spatial co-ordinates. This Californian research explored SOM mapping techniques relating to wine flavors. The study examined the similarities and differences in wine varieties using 85 flavors from seven wines, all from the United States of America, in particular the California region. The research concluded that SOM techniques enable analysts to determine wine taste and smell correlations that originate from the grape “cultivars” and factor dependencies between climate, soil, winemaking techniques and the ageing process.

By studying the dark lines in the SOM, the main classes among the wines analyzed could be established. For instance, the dark line at the bottom left of the SOM (figure 7) indicates that Chardonnay is up to 85% different from the rest of the wines based on its flavor properties of fruit, spice, soil. There is also a boundary between Pinot Noir and Merlot. Meanwhile, Riesling, Sauvignon Blanc, Syrah and Cabernet Sauvignon indicate the most similar qualities. The properties used in the research were obtained from website information collated using Google search engine for these wine varieties.

So we conclude that this analytical tool is specifically appropriate as one of the collaborating technologies useful for our research.

The description of Viscovery as given by its developers says that, Kohonen nets are (artificial) neural networks which adapt themselves in response to input signals and on the basis of the Kohonen algorithm. They consist of nodes, which are spread uniformly on a grid and are “functionally” connected to their neighboring nodes. In most cases, two dimensional grids are used, however, Kohonen nets with one-dimensional or multi-dimensional grids are possible as well.

Viscovery SOMine is a tool for advanced analysis and monitoring of numerical data sets. Based on the concept of Self-Organizing Maps (SOM), it employs a robust variant of unsupervised neural networks, namely Kohonen’s Batch-SOM, which is further enhanced with a new scaling technique for speeding up the learning process. Within Viscovery® SOMine software, the SOM is realized by a two-dimensional hexagonal grid. Starting from a set of numerical, multivariate data, the “nodes” on the grid gradually adapt to the intrinsic shape of the data distribution. Since the order on the grid reflects the neighborhood within the data, features of the data distribution can be read off from the emerging landscape on the grid. The training process is time-consuming and may be followed by the user on a progress graph. In a second step, the data representation contained in the trained SOM is systematically converted for use in a spectrum of visualization techniques.

Evaluating dependencies between components, investigating geometric properties of the data distribution, searching for clusters, monitoring new data, just to mention a few options, thereby become an intuitive and inspiring interactive process. In parallel, numerical information, such as cluster statistics, can be retrieved on demand at any time.

Attaching a specific meaning to a particular map area is a very useful feature of Viscovery® SOMine. This functionality enables the placement of descriptive labels at any node in the map [15].
The SOM is composed by two layers (figure 8). The input layer has the vector inputs $\mathbf{x}(t) \in \mathbb{R}^m$, with each component $k (1 \leq k \leq m)$.

Each input node is connected to all nodes of output layer with a weight denoted $w_{ijk}$.

In the training process the weights of the maps change based on the following expression,

$$w_{ijk}(t+1) = w_{ijk}(t) + \alpha(t) h_{ij}(t) - g_{i}(t) \cdot (x_{k}(t) - w_{ijk}(t))$$

Where $\alpha(t)$ is the learning rate and $h(\cdot)$ is the neighborhood distance and it depends of $i$ (the node $(i,j)$) and $g$ (the winner node). The winner node $g$ is calculated using the Euclidean distance between the input vector and the weight vector of each node.

7. GIS TECHNOLOGY

Geographic Information Systems (GIS) have enabled significant advances in many fields of application including local authority service planning and provision, agricultural management and demographic analysis [16]. This technology is used in any context where spatially referenced data is available and can be related to still or moving images, maps and other numeric or non-numeric data. It is appropriate to say now after only nearly two decades since their inception, that GIS are ubiquitous facets of many aspects of corporate and public management, and the sciences of Geography, Geodesy and Computer Science. The integration of remote sensing techniques, cadastral and other spatially referenced data, maps and other images, together with methods of analysis and extrapolation of data element relationship and inter-dependencies from multivariate analysis and other statistical forms of combinatorial correlations, has brought about the evolution of a recent field of research and teaching known as Geomatics or Geo-informatics. This field is in itself collaborative and trans-disciplinary [17].

Three GIS products are being used in the research but this one is DIVA (www.divagis.org), which was developed by NASA and collaborators specifically for climate and related agricultural influence data. (See also [18]).

8. EXAMPLES OF DATA COLLECTED

The data collected and analyzed in real time so far in this early stage of the project is only that relating to climatic influences. Some data relating to sap rise in vines has been collected also and that was described in section 5 of this paper. If we combine these two sets of data and also include some historical data from a source provided by the Chilean Government, we can see how these are ingested into the database of variables that has been designed and developed so far. The database design is depicted below, together with a table of variables against which the vector data we are collecting is classified.
9. CONCLUSIONS

It is of course, too early to reach any significant conclusions relating to actual data combinations and the production of scenarios or perhaps predictions. It is clear though, that the concept of the research project has captured the imagination of numerous scientists and industry partners who see the potential for the work and who would find the outcome both useful in practice and also regard it as a contribution to science.

Some upcoming areas of investigation in this research include remote sensing using satellite imagery and image processing techniques for cloud cover classification as demonstrated in [19, 20] and entry of resulting values to the neural network synthesizer. We are conducting trend analysis for climate data in combination with atmospheric and other plant growth and condition data. It is intended to juxtapose these results with for example, the concept of Thermal Sensation, which in human physiology relates to the effects of temperature changes to the body, especially in extremes of cold and heat [21]. We expect to test a similar set of variables and parameters used to characterize heat alterations in humans against plant conditions that vary in temperature and other climatic influences. Sun, wind, rain and temperature variations are obvious influences on plant condition and it is expected that human receptors, such as fingers, toes, eyes and others that characterize the extremes of thermal sensitivity change, will correlate one set to another.

Another temperature related area of research is Thermal Sensitivity Analysis, which has been used in materials science for measuring the effects of temperature variations in metals in [22] and some work has been undertaken with this analytical approach to measure the effects of heat transfer in generator rotors. The generalization of the equations that have been developed for this area of investigation have also been used for estimating the influence of heat variation in plants [23].

In the project being described in this paper, we are attempting to apply a similar variable set to vines in terms of what we can actually measure as reliable influence data.

Hyperspectral Remote Sensing is yet another area of related investigation [24]. The sensors used in this field, which is also known as imaging spectroscopy, are being developed to detect and identify minerals, terrestrial vegetation, as well as synthetic materials. We contend that this area of science should not be ignored as potentially contributing to the comprehensiveness of the data set we are building to reflect a broad spectrum of environmental influences on crop production.

The work continues and in essence will grow as a project that exemplifies the notion of collaborative technologies in action as they integrate with the human context for international research collaboration.

REFERENCES


