This page is for indexing purposes and will not be printed in the conference proceedings book.

The text area is 29 × 45 picas (~12.3 cm × 19 cm) to fit in the printed proceedings book without reduction. When printed on letter paper there will be wide margins.

Author(s)

First Name	Middle Name	Surname	Role
Flávio	R. da S.	Oliveira	First Author

Affiliation

Organization	URL	Email
Department of Computing Systems Polytechnic School of Engineering - Pernambuco State University, Rua Benfica, 455 - Bloco "C" - 3o. Andar CEP: 50.720-001 – Recife/Pernambuco, Brazil	www.dsc.upe.br	flaros@hotmail.com

Author(s)

First Name	Middle Name	Surname	Role
Diogo	F.	Pacheco	Second Author

Affiliation

Organization	URL	Email
Department of Computing Systems Polytechnic School of Engineering - Pernambuco State University, Rua Benfica, 455 - Bloco "C" - 3o. Andar CEP: 50.720-001 – Recife/Pernambuco, Brazil	www.dsc.upe.br	diogofpacheco@yahoo.com.br

Author(s)

First Name	Middle Name	Surname	Role
Amanda		Leonel	Third Author

Affiliation

Organization	URL	Email
Department of Computing Systems Polytechnic School of Engineering - Pernambuco State University, Rua Benfica, 455 - Bloco "C" - 3o. Andar CEP: 50.720-001 – Recife/Pernambuco, Brazil	www.dsc.upe.br	amandaupe@yahoo.com.br

Author(s)

First Name	Middle Name	Surname	Role
Fernando	B. de	Lima Neto	Corresponding Author

Affiliation

Organization	URL	Email
Department of Computing Systems Polytechnic School of Engineering - Pernambuco State University, Rua Benfica, 455 - Bloco "C" - 3o. Andar CEP: 50.720-001 – Recife/Pernambuco, Brazil	www.fbln.pro.br	fbln@dsc.upe.br

Publication Information

Pub ID	Pub Name	Pub Date
ASABE will complete	*Computers in Agriculture 2006* (24-26 July, Lake Buena Vista, Florida, USA).	Pub. Date July 2006

Intelligent Support Decision in Sugarcane Harvest

Flávio Rosendo da Silva Oliveira[1], Diogo Ferreira Pacheco[1], Amanda Leonel[1], Fernando Buarque de Lima Neto[1,2]

1) Department of Computing Systems – Polytechnic School of Engineering (*Escola Politécnica de Pernambuco*)
Pernambuco State University. Rua Benfica, 455 - Bloco "C" – 3o. Andar * CEP: 50.720-001. Recife / Pernambuco, Brazil.
2) Corresponding Author: fbln@dsc.upe.br

Abstract. This paper presents a computing approach to support harvest decisions in sugarcane utilizing artificial intelligence (AI). The proposed two-step Decision Support System (DSS) starts with an AI technique called Artificial Neural Network (ANN), which is utilized to forecast agronomical performance indicators. Next, an heuristic that uses Genetic Algorithms (GA) is applied to search, and then recommend, suitable areas to be harvested. This work includes some experiments with real data where the ideas put forward are tested. The results of these experiments proved our approach: (i) useful to decision makers and (ii) easily coupled to current Management Information Systems (MIS) existing in most sugarcane mills.

Keywords. Sugarcane, Harvest, Decision Support Systems, Artificial intelligence, Artificial neural networks. Genetic algorithms.

Introduction

Sugarcane is one of the major sources of carbohydrates for human feeding. More recently, sugarcane is also used as renewable fuel. In Countries such as Brazil, for example, ethanol extracted from sugarcane represents 20-30% of the fuel that run the 23 million vehicles of the Brazilian fleet. All that increasing energy demand requires more productivity in the fields.

One can achieve greater productivity by using computers for supporting harvest decisions; previous works proved this possible [Lima Neto, 1998] [Trigo *et al*, 2005] [Pacheco *et al*, 2005]. The rationale for this work is quite simple: to couple mills demands with acceptable agronomical productivity levels prior to harvesting, which is also the objective of this work. *I.e.*, a computer system that helps on harvest decision for sugar-cane, by bringing together intelligent computer techniques and structured decision dialogues.

In this work approach artificial neural network is utilized for forecasting agronomical performance indicators, followed by a customized decision support system. This DSS utilizes the predictive agronomical information produced by the artificial neural networks to, together with genetic algorithms principles, couple predictions to the daily needs of sugarcane for milling. As a result, ANN and GA embedded in a custom made DSS make it a useful tool for providing the decision maker with the necessary information to better administer the harvest (*i.e.* to manager and to control it).

Some results obtained from the DSS proposed and presented here are in the end of this paper. The reader can then evaluate how better decisions and, consequently, better productivity results can be achieved (*i.e.* selection of plots for harvesting at their agricultural peaks).

Background

This work is heavily grounded in the field of Artificial Intelligence. In this section four short reviews on the key concepts utilized here will be carried out: (i) sugarcane productivity factors and indicators, (ii) decision support system, (iii) artificial neural network and (iv) genetic algorithms.

Sugarcane productivity factors and indicators

As in any other business, management information systems of sugarcane mills are prolific in having historical data regarding *factors* and *indicators* related to production. By *factors* we mean all aspects that may cause relevant impact on productivity levels; by *indicators* we mean ways to check (good/bad) results of the production process as a whole [Lima Neto, 1997].

To test our model, we used real (historical) data of three sequential years of a Brazilian sugarcane mill of the Southwestern State of São Paulo [Lima Neto, 1998]. After a careful analysis on the data and talks to Agronomists, we chose the following productivity factor from the data-set (for each plot): sugarcane type, topography, sowing date, soil type, age (regarding harvest day) and number of cultivation cycles. As productivity indicators, *i.e.* variables to access quality of future decisions, the choices were (in previous years): (i) sucrose – PCC, (ii) vegetal fibers levels and (iii) biomass per squared area – TCH; these three, in relation to same productivity factors (*i.e.* average values for the same plot within annual productions).

Following dataset selection, as this paper explains later, productivity factors and indicators were used to train the intelligent technique: Artificial Neural networks. Obviously that another variable set could be used in the modeling process. Most likely it would evoke different results as ANN, like most AI techniques, are non-deterministic.

Decision Support Systems (DSS)

DSSs are mainly used in mid-managerial levels of organizations in order to help semi-structured decision processes to be made. It is common to find DSSs linked with Management Information Systems (MIS). MISs, in that case would function as data sources to subsidize the decision process supported by DSSs. In this sense, the combination of MIS and DSS result in a powerful administrative tool, granting two distinct and highly complimentary functionalities: control and management, respectively. By semi-structured decision we mean decisions that may change in some form thru time. Thus, DSSs are very useful to help decision makers to analyze possible scenarios and the effects of their decisions upon those scenarios.

DSSs should be fast enough to produce various scenarios analyses within small amounts of time as well as to be flexible for interacting with the decision maker in a non-monotonic manner. A DSS has to have flexibility to incorporate some businesses changing aspects, for example: priorities of the day during harvest. So, DSSs have to combine information extracted from the company's database as well as to profit from the expertise of the decision maker during the interaction (*i.e.* decision) process.

Based on the myriad of mathematical and analytical models in existence, as well as problem orientated applications of DSSs it is worth mentioning some of their use categories [Sprague, 1996][Turban, 1995]:

1. Simulations – most commonly used and simple types of DSSs. DSSs for this kind of problem have plenty of "what – if" questions, repeatedly posed throughout the decision process in order to determine the impact of changes or output of given variables (*i.e.* productivity factors). This kind of DSS, normally, has an interaction that allow users to visualize the impact of their choices on other variables or parts of the system (*i.e.* decision system);
2. Reach goals – medium complexity types of DSSs. They start from pre-set goals to whose the system recommends main and alternative routes. They also may include effort assessing functionalities;
3. Optimizations – aim at specific targets by playing with variables of an analytical model of the problem. They are the most advanced type of DSS and not for chance, the most complex to build.

Artificial Neural Networks

In the 1940 decade, the *organization* of human brain inspired scientists to build the first computers as well as it inspired the ground breaking research that unveils Artificial Intelligence. However, only forty years later, in mid-1980, the human brain *functioning* has became an effective paradigm for intelligent applications to problems in industry.

Over the past twenty years AI research unfolded in many sub-areas and techniques. Artificial neural networks, a leading ones of them, is successfully used for forecasting, exactly because their biological inspiration (as a chief ability of the brain is precisely to compute outcomes of facts and actions).

Artificial neural networks encompass distinct algorithms, varying in internal organization, computing power and applicability. Among their diversity, ANNs share in common: learning ability, distributed computation and generalization capability. Haykin (1994) defines ANNs as a set of distributed processors, highly parallel in nature, all of them with the distinctive ability to acquire, store and make knowledge available for future uses. Similarly to the human brain, ANNs are composed of (artificial) neurons – processing units – and synapses, the latter loci of learning storage. Together they grant to ANNs most of the features described above.

Multilayer Perceptrons – MLP [McClelland, 1986] was the ANN type we utilized in this work to produce inferences of (future values of) productivity indicators. All MLP architectures used had three layers of processing units. Differently from other computing applications, ANNs need to be trained with patterns (*i.e.* historical data instances) that inform aspects of problem being modeled. Along the ANN training phase, after presentation of each pattern, the learning algorithm adjust weights of neurons (*i.e.* changes synaptic values) by comparing differences of computed output to the desired value obtained from training data.

Details of training, dataset and validation processes can be found in previous works of the authors of this paper [Lima Neto, 1998] [Trigo *et al*, 2005] [Pacheco *et al*, 2005].

Genetic Algorithm

Genetic algorithm is an optimization and search (another intelligent) technique based on principles of genetics and natural selection. In this technique specific features of problems to be solved are coded in genes, which comprise an individual (*i.e.* chromosomes) as all genes assume single values for every feature coded. GA allows a population composed of various individuals to evolve under specified selection rules to a state that maximizes its fitness [Haupt, 2004].

When applying GA for solving a problem, the first step is to make a suitable representation to describe the problem states. The most common way is to use a string of bits. It also must be selected an objective function, so called fitness function, to evaluate the "merit" of each candidate solution [Tzung-Pei, 2002].

Some characteristics of GA are highlighted below (all of them highly relevant to the present work):

- Deals with a large number of variables
- Optimizes variables with extremely complex cost surfaces
- Can avoid local maxima or minima
- Can provide a list of optimal variables, instead of only one
- Simultaneously searches a large sample of the cost surface

Figure 1 illustrates the complete processing cycle of a simple GA (same as the one utilized here). Each box in the figure is a set of computing tasks that reproduces evolution. Here "evolution" is a metaphor for searching of a suitable solution for the posed decision problem.

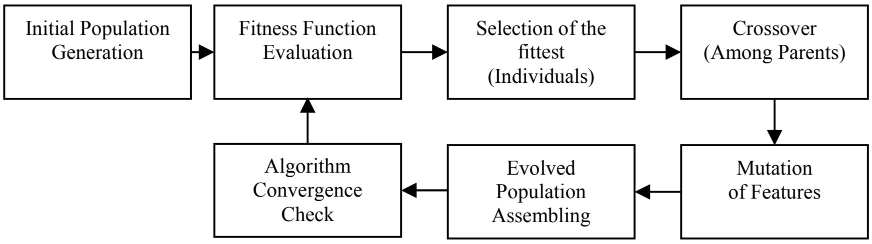

Figure 1: Processing cycle for a simple GA

At the end of each processing cycle the convergence check is performed to assess whether the algorithm reached its objective; if so, it can stop. Otherwise, it must continue the search by evolving the population (of candidate solutions). In this work, the GA technique was used to find the most suitable areas (plots) to be harvested. The search was based on predictive agronomical information produced by the ANNs and on some decision criteria, provided by the decision maker.

Abstract Model to Decision Making

The decision making concerning the harvest of sugarcane plots is complex because of many reasons, namely, the number of variables involved and the need to couple it with conflicting interests. For example, it is expected for the decision maker to select sugarcane plots for harvest with high sucrose content as well as high vegetal fiber level. The first, for obvious economical reasons and the second for attending energy demands of sugarcane mills; sometimes sucrose and fibers levels cannot be maximized simultaneously. In the present work, an abstract model to decision support was developed in order to tackle with the complexity of the task and the need of formal models to deal with this problem.

Our approach deals with the problem in a broad and consistent way, as we start with a theoretical basis before implementing the decision support system (DSS) itself. By using an abstract model we aim to facilitate: (i) evaluating the impact of decision variables in the problem and (ii) adapting the solution to different application set-up. Both are desirable features because they allow effective solution customization.

The formalism for the abstract decision model put forward here is: -given a decision problem $\mathbf{p_k}$ from a class of problems $\mathbf{P}$, a decision maker $\mathbf{m_k}$ from a class of decision makers $\mathbf{M}$, where $\mathbf{m_k}$ is necessarily empowered to solve $\mathbf{p_k}$. Assuming the existence of a decision space for problems of class $\mathbf{P}$, $\mathbf{D_P}$ so that $\mathbf{D_P} = \{d_1, d_2, ..., d_n\}$. Then every $\mathbf{d_k}$ is a decision, potentially capable, of solving problems of class $\mathbf{P}$. To solve an specific problem $\mathbf{p_k}$ each candidate decision $\mathbf{d_k}$ ought to encompasses a subset of decision components $\mathbf{C}$, upon which $\mathbf{m_k}$ should reason separately. In this way, $\mathbf{d_k} = \{c_1, c_2, ..., c_n\}$, where every $\mathbf{c_j}$ belongs to $\mathbf{C}$. Still, each decision component $\mathbf{c_j}$ is composed by a set of attributes which characterize it, so that $\mathbf{c_j} = (a_1, a_2, ..., a_n)$, check Figure 2 for an overview of this formalism. Ranges of values for every $\mathbf{a_i}$ are drawn from the **M**-**P** interaction and can be either qualitative or quantitative.

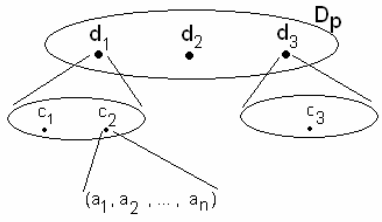

Figure 2: Decision space over Problem class **P** encompassing decisions, components and their attributes

To establish priorities among candidate decisions $\mathbf{d_k}$ of $\mathbf{D_P}$, all able to solve $\mathbf{p_k}$, and also to assess their suitability regarding the posed problem, we devise an *ad hoc* metric, *i.e.* a function that maps attributes of components of a decision: $\mathbf{f_i\ (a_i)}$. Output values of $\mathbf{f_i}$ are normalized within the interval [0,1] to avoid distortions during decisions evaluation. For example, scale differences of attributes could bias some decisions.

For the metric to work fine, $\mathbf{m_k}$ must provide a weight vector W = $\{w_1,w_2,...,w_n\}$; $\mathbf{w_i}$ should be individually ascribed to each attribute of decision's components. Finally, another vector V = $\{v_1,v_2,...,v_n\}$ is the validity criterion concerning every attribute. This vector should also be provided prior to the decision.

Using the weights vector W, it is possible to evaluate the relevance of each component $\mathbf{c_j}$ to the decision d_k, which is given by Equation (1):

$$R(C_j) = \frac{\sum_{i=0}^{n} w_i * f_i(a_i)}{\sum_{i=0}^{n} w_i} \tag{1}$$

Finally, we proposed that a good decision ought to: (a) maximize all associated return criteria of each decision components (as well as to be valid according to **V**) or (b) minimize the resulting error from expected results for the problem at hand (also according to **V**). Put in this way, the overall decision process is reduced to maximizing/minimizing the function $\mathbf{F(d_k)}$, which is presented in two forms, according to the case (see Equation 2 and Equation 3, respectively).

Maximizing Return
$$F(d_k) = \sum_{j=0}^{n} R(C_j) \tag{2}$$

Minimizing Error
$$F(d_k) = \sum_{j=0}^{n} \frac{1}{R(C_j)} \tag{3}$$

The approach proposed here is to use GA to search for the decision $\mathbf{d_k}$ in the decision space $\mathbf{D_p}$ which maximizes/Minimizes the function $\mathbf{F(d_k)}$ and which is valid, according to the validity criteria set **V**.

Applying the Model to Decision in Sugarcane Harvest

To test the ideas put forward, an especial DSS was implemented – *Genetic Decision Support System* (gDSS). When applied to sugarcane harvesting, the new system aims at maximizing the agronomical return of a set of sugarcane plots. According to the developed model:

1. The decision space $\mathbf{D_p}$ is composed by different sets of plots comprising a days harvest
2. Each decision component is a plot, represented by a tuple: Plot-ID, PCC, TCH, Fiber and Plot-squared area (remember that PCC, TCH and Fiber were provided by ANN forecasting).
3. Two functions were defined to evaluate the attributes PCC and Fiber:

$$f_{pcc}(PCC_i) = \frac{TCH_i * Area_i * PCC_i}{MAX\ (TCH * Area * PCC)} \tag{4}$$

$$f_{Fiber}(Fiber_i) = \frac{TCH_i * Area_i * Fiber_i}{MAX\ (TCH * Area * Fiber)} \tag{5}$$

4. The decision maker informs weights he judges for each attribute W={w_{fiber},w_{pcc}} and validity criteria, V = {*Desired ton of sugarcane*, *Maximum Area*, *Minimum PCC*, *Minimum Fiber*}
5. Based on Equation (4) and Equation (5), now a function was defined to evaluate the relevance of a component relating to PCC and Fiber, simultaneously:

$$R(C_j) = \frac{f_{Fiber}(Fiber_i) * w_{Fiber} + f_{pcc}(PCC_i) * w_{PCC}}{w_{Fiber} + w_{PCC}} \quad (6)$$

6. The decision making process, was considered then as finding the plots set which maximizes the function **F(d$_k$)**:

$$F(d_k) = \sum_{j=0}^{n} R(c_j) \quad (7)$$

Simulation and Results

In order to test the model, a set of 418 plots were used to simulating a real sugarcane mill decision environment. The combinatorial explosion of factors involved in this search for the best solution would not be viable barehanded. That is why a DSS such as the one presented here should be applied. In gDSS, GA was utilized to find a suitable solution, according to the parameters provided by the decision maker.

The encoding used was similar to that used to solve the Knapsack Problem. The chromosome is a string of bits as long as the number of plots available to harvest. Each position encodes only if the related plot is part or not of a given decision **d$_k$**. The fitness function is composed primarily of **F(d$_k$)** which sums-up the relevance of each plot in the computed decision **d$_k$**. For each validity criterion unattended, the fitness value is decreased cumulatively by 50%.

During the selection phase of the GA processing cycle, the chromosomes are ordered regarding their fitness. Then, we used the heuristic of copying 10% of the fittest chromosomes directly to next generation. The other chromosomes of the next generation were obtained, similarly to what is observable in basic genetics: by crossover (*i.e.* pairing and combining the genes of parent chromosomes). This process carries on until a new complete population is obtained. The evolutionary cycle stops when either a maximum number of generations are evolved or sought sugarcane productivity objective is achieved. In each simulation performed here we used a population with 50 chromosomes and 50 evolutions.

To validate our tool, it was defined a plausible scenario whose validation criteria (provided by the decision maker) were: *Minimum PCC = 16, Minimum Fiber = 15, Desired TCH = 650, Maximum Area = 10 plots.* Additionally, the weights used among factors were $w_{pcc} = 10$ *e* $w_{fiber} = 5$. The sugarcane productivity indicators were generated by a *neural engine* (based on ANNs) that was developed by the authors prior to this work. The generated file contains predictive information of 418 plots, namely, plot identification, PCC, TCH and Fiber. Results of the gDSS – decision method introduced here – are compared to other methods over the same 418 plots. The two other methods were: (i) simple manual selection and (ii) another DSS developed by Pacheco (2006); both use linear programming. In order to simplify the comparisons of method, all plots were considered as having same areas; see results in Table 1 and Table 2.

Table 1. Overall Performance Comparison among distinct Decision Methods

Method	**Plots-ID (Recommended for Harvesting)**	**TCH**	**Avg. Fiber**	**Avg. PCC**
Manual	34, 56, 102, 169, 199, 238, 365, 385, 404	649.0026	15.8376	16.5478
[Pacheco, 2006]	26, 34, 56, 102, 131, 169, 199, 365, 385, 404	667.0466	15.8349	16.5431
gDSS (this work)	22, 314, 290, 335, 194, 147	649.8212	15.1324	16.1012

Table 2. Productivity indicators of Recommended Plots

Plot-ID	**TCH**	**Fiber**	**PCC**
22	92.5419	14.7150	16.5217
147	126.3737	15.0594	15.4002
194	108.8569	15.6233	15.9904
290	83.552	15.6494	16.1731
314	120.9601	14.9791	16.4559
335	117.8333	14.7683	16.0659
	Sum. TCH	**Avg. Fiber**	**Avg. PCC**
	649.8212	15.1324	16.1012

Conclusion

The selection of plots during sugarcane harvest involves many and non-trivial decisions to be taken. This work has contributed with a novel type of DSS that utilizes genetic algorithms to help on the search for a suitable selection. As decision criteria are varied and dependent on the problem as well as on the decision maker, the contributed gDSS is equipped to allow users to set it up towards better recommendations.

When compared to other decision methods, namely, (i) Manual and (ii) Pacheco's [Pacheco, 2006], gDSS performed well recommending plots that, when considered as a whole, produce average productivity indicators. However, the number of plots recommended by gDSS was consistently smaller than the other two decision methods, which can be an operational advantage. Additionally, an automatic recommending system such as the gDSS can be a tactical advantage to the agricultural business. As it can speed-up the decision process as well as help to avoid common errors such as misinterpretation of data or forgetting.

As gDSS utilizes GA it become non-deterministic, this can be another tactical advantage as the system can provide new sets of recommended plots (if more evolutionary steps are allowed). This feature is not observed on the other two linear methods gDSS was compared to.

We conclude by highlighting that gDSS is an useful tool to decision makers and it is simple to couple them with existing MIS. At the moment, other logistic aspects of the decisions are now being incorporated to increase suitability of recommendations.

References

Diogo Pacheco, Filipe Regueira and Fernando Buarque de Lima Neto. Using Artificial Neural Networks on Sugarcane Harvest to Predict Sucrose, Fiber and Productivity (in Portuguese), in *ALCOOLbrás*. São Paulo, Brazil, v. 90, pp. 60-63, 2005.

Diogo Pacheco. *An Intelligent Decision Support System for Agriculture Harvest* (in Portuguese), Technical Report presented as Graduation Monography to Department of Computing Systems – Polytechnic School of Engineering – Pernambuco State University, 2006.

Efrain Turban. *Decision Support Systems and Expert Systems*, 4th. Edition , Prentice-Hall International Editions. New Jersey, USA, 1995.

Fernando Buarque de Lima Neto and Teresa Bernarda Ludermir. Decision Support on Sugarcane Harvest (in Portuguese), in *IV Brazilian Symposium on Neural Networks – SBRN*. Goiania, Brazil, 1997.

Fernando Buarque de Lima Neto. *Managerial Decision Support, based on Artificial Neural Networks* (in Portuguese), Master Dissertation presented to Department of Informatics, Federal University of Pernambuco, Recife, Brazil, 1998.

James McClelland, David Rumelhart, and Geoffrey Hinton. The Appeal of Parallel Distributed Processing, in D. Rumelhart, J. McClelland and PDP Research Group (Editors). *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, MA: MIT Press, 1986.

Randy L. Haupt and Sue E. Haupt. *Practical Genetic Algorithms*. 2nd ed. Wiley-Interscience, 2004.

Simon Haykin. *Neural Networks – A Comprehensive Foundation*. Prentice-Hall International Editions. New Jersey, USA, 1994.

Ralph Sprague Jr. and Hugh J. Watson. *Decision Support for Management.* Prentice-Hall International Editions. New Jersey, USA, 1996.

Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach.* 2nd Edition. Prentice-Hall International Editions. New Jersey, USA, 2003.

Tiago Trigo, Paulo Batista Jr. and Fernando Buarque de Lima Neto. Artificial Neural Networks on Sugarcane Harvest (in Portuguese), in *Proceeding of V Brazilian Congress of Agroinformatics*. Londrina, Brazil, 2005.

Tzung-Pei Hong, Ke-Yuan Huang, Wen-Yang Lin: Applying genetic algorithms to game search trees. Soft Comput. v. 6(3-4) pp. 277-283, 2002.