Mining Time Series of Meteorological Variables Using Rough Sets – A Case Study

C. Fernández-Baizán[1], **C. Pérez-Llera**[2], **R. Portaencasa**[3], **J. Feito**[4]

Abstract. In this paper, a rough sets based approach to short-term prediction of one variable -wind speed- is proposed. The inputs to the system are basically historical values of the predicted variable as well as a number of other support variables. This article attempts to deal with the problem of time within the framework of rough sets. The rough set theory has emphasized the reduction of information necessary to acquire desired knowledge. This is particularly important when we are dealing with time. Several steps must be taken to achieve our objective, which are presented here as work guidelines.
This work is supported by a grant between the University of Oviedo (Oviedo, Spain) and the enterprise European Bulk Handling Installation S.A. (Gijón, Spain) under the Life Environmental Research Project of the European Union.

1 INTRODUCTION

The Dry Bulk Transshipment Terminal in the Port of Gijón-Spain, European Bulk Handling Installation S.A. (EBHISA) offers facilities for the unloading, storage and transshipment of any kind of bulk cargoes. The terminal has 652 meters long quay.

Discharge has achieved rates up to 70.000 tonnes, unloading vessels of 150.000 m/t dwt. The Terminal can accommodate two vessels of up 200.000 dwt at a time, and operates 24 hours a day seven days a week, with three shifts unloading some 11 million tonnes per year.

The terminal ha two 50 tonne grab unloaders of 2400 tph each and two stacker reclaimers providing a capacity of 2800 tph.. The Terminal also includes a loader for 35000 dwt vessels with a 2600 tph cargo capacity.

There is storage area of 150394 m2 at Minerales quay with an approximate capacity of 600000m/t at open area.

All cargo movements are controlled by the Terminal's Control Room. EBHISA takes special care to ensure that the operations doesn't contaminate the atmosphere.

In order to avoid the environmental contamination produced during the load/unload operations of goods (mainly coal), accurate forecasts of winds are crucial. Until now forecasting methods rely heavily on the knowledge and experience of a local meteorologist.

Rough Sets methodology is applied in an expert system setting to construct a meteorological forecasting tool which predicts the local short-term wind speed in the Port of Gijón (Spain). The expert system is designed in order to discharge the watering and prevent the formation of big clouds of powder.

Due to the numeric character of the data, a statistical analysis and a study of neural nets of temporal series generated with part of this data has been carried out in [11], [12].

One specific problem relating to many interacting events, is the problem of time dependencies. Traditional rough set theory has emphasized the reduction of information necessary to acquire desired knowledge. Therefore in this paper, we will focus on objects and attributes that are dependent on each other in time. We will expose a framework that extends the traditional information system of rough sets to include a sequence of events.

Some experiments have been carried out on each 10 min. meteorological data.

Knowledge of rough set theory is required [7], [8].

2 RELATED RESEARCH

2.1 Rough Sets Theory

The original Rough Set model was proposed by Pawlak [7]. This model is concerned with the analysis of deterministic data dependencies. According to Ziarko [16] Rough Set Theory is the discovery representation and analysis of data regularities. In this model, the objects are classified into indiscernibility classes based on pairs (attribute, values).

The following are the basic concepts of the Rough Set model.

Let *OB* be a non-empty set called the universe, and let *R* be an equivalence relation over the universe *OB*, called an

[1] Fac. Informática, Univ. Politéc. de Madrid, email: cfbaizan@fi.upm.es
[2] ETS Ingenieros Informáticos, Univ. Oviedo, email: cpllera@etsiig.uniovi.es
[3] ETS Ingenieros Telecomunicaciones, Univ. Politéc. de Madrid
[4] Fac. Ciencias Económicas, Univ. León, email: ddejfg@unileon.es

Indiscernibility Relation which represents a classification of the universe into classes of objects which are indiscernible or identical in terms of the knowledge provided by the given attributes. The main notion in Rough Sets Theory is that of **Boolean Approximation Space** which is formally defined as:

$$A = (OB, R) \qquad (1)$$

Equivalence classes of the relation are also named elementary sets. Any finite union of elementary sets is called a definable set. Let's take $X \subseteq OB$ which represents a concept. It is not always the case that X can be defined exactly as the union of some elementary sets. That's why two new sets are defined: $\underline{R}(X) = \{\ o \in OB \mid [o] \subseteq X\ \}$will be called **Lower Approximation** approximation $\overline{R}(X) = \{\ o \in OB \mid [o] \cap X \neq \emptyset\}$ will be called **Upper Approximation**. Any set defined in terms of its lower and upper approximations is called a **Rough Set**.

The main computational effort in the process of data in rough sets theory is associated with the determination of attribute relationships in information systems. An **Information System (IS)** is a quadruple: $S = (OB, AT, V, f)$ where:

- OB is a finite set of objects
- AT is a finite set of attributes
- $V = \cup V_{a \in \mathrm{AT}}$ being V_a the values of attribute a
- f: $OB \times AT \rightarrow V$ is a total function such that $f(o,a) \in V_a$ for every $a \in AT$, $o \in OB$, called **information function**.

Let $Y = \{\ Y_1, Y_2, \ldots, Y_n\ \}$ be a **classification** (partition) of OB in S, and $P \subseteq AT$. Subsets Y_j ($j = 1, 2, \ldots, n$) are classes of Y. Then, $\underline{P}(Y) = \{\ \underline{P}(Y_1), \underline{P}(Y_2), \ldots, \underline{P}(Y_n)\ \}$ and $\overline{P}(Y) = \{\ \overline{P}(Y_1), \overline{P}(Y_2), \ldots, \overline{P}(Y_n)\}$ are called the ***P*-lower** and the ***P*-upper approximation** of classification Y, respectively. $\underline{P}(Y)$ and $\overline{P}(Y)$ are also called ***P*-positive region** and ***P*-negative region** of Y. $BNP(Y) = \overline{P}(Y) - \underline{P}(Y)$ is called ***P*-boundary** or **P-doubtful region** of Y.

The **accuracy of classification** is defined as

$$\alpha_P(Y) = \sum_j card(\underline{P}(Y_j)) \,/\, \sum_j card(\overline{P}(Y_j)),$$

and expresses the percentage of possible correct decisions when classifying objects employing the knowldege P. Inexactness of a decision is due to the existence of a doubtful region. The accuracy coefficient expresses how large the boundary region is. The accuracy of approximation of Y by a set P of attributes is a measure of vagueness and captures how accurately the set Y_i is definable in Y based on knowledge P. Accuracy measure is intended to capture the degree of completeness of our knowledge about the classification Y.

The coefficient,

$$\gamma_P(Y) = \sum_j card(\underline{P}(Y_j)) \,/\, card(OB)$$

is called the **quality of approximation** of Y by P or, in short, **quality of classification**. If the quality measure is 1 then the information system is deterministic, if less than 1 it is roughly deterministic, and if 0 it is totally non-deterministic. It expresses the percentage of objects which can be correctly classified to classes of Y employing knowledge P.

Core and reducts are two fundamental concepts of rough sets. A reduct is the essential part of an IS which can discern all objetcs discernible by the original IS. A core is the common parts of all reducts. The set of attributes $R \subseteq Q$ **depends** on the set of attributes $P \subseteq Q$ in S (denotation $P \rightarrow R$) iff $IND(P) \subseteq IND(R)$. Discovering dependencies between attributes is of primary importance in the rough sets approach to data analysis. Another important issue is that of **attribute reduction**, in such a way that the reduced set of attributes provides the same quality of classification as the original set of attributes. The minimal subset $R \subseteq P \subseteq Q$ such that $\gamma_P(Y) = \gamma_R(Y)$ is called ***Y*–reduct** of P (or, simply, **reduct** if there is no ambiguity in the understanding of Y). Y–reduct of Q is also called **minimal set** (or **subset**) in S. Let us notice that an information system may have more than one Y–reduct. RED_Y (P) is the family of all Y–reducts of P. Intersection of all Y–reducts of P is called ***Y*–core** of P, i.e. $\mathrm{CORE}_Y (\mathrm{P}) = \cap\, \mathrm{RED}_Y (\mathrm{P})$. The only Y–core of Q is the set of the most characteristic attributes wich cannot be eliminated from S without decreasing the quality of approximation Y.

Formally, a **Decision Table** S is a quadruple $S = (OB, C, D, V, f)$. All the concepts are defined similarly to those of information systems; the only difference is that the set of attributes has been divided into two sets, C and D, which are conditions and decision respectively.

2.2 Temporal Reasoning

Much work has been carried out with mathematical models based on e.g. Fourier series and probability distributions and correlations [5].

Interesting paths of research have also been taken in the area of temporal logic. There have been attempts to model real-time sequences with temporal logic. This was suggested by Jonathan Ostroff [6], where he introduced the framework of Real-Time Temporal Logic (RTTL). Within this framework, he used an event variable, n, to refer to events, and a clock variable, t, to refer to the clock value. This gave him the opportunity to state more than just the proper sequence of events.

2.2 Temporal Reasoning With Rough Sets

One approach is reported in [1]. Market data from the Hughes Research Laboratory have been analyzed, and some success in predictions has been reported. Dynamic reducts are utilized, and have been reported as a necessity in this application. The data used variants of each indicator representing both the present level and previous trends.

Another approach is reported in [3], [4]. They have based their analysis on Canadian stock market data on a monthly basis. No predictions are reported quantitatively however. The translation from a time series to a traditional information system is introduced in this paper.

Time varying information systems have also been discussed in [9]. We have not found references in applications of Rough Sets to meteorogical prediction.

3 ROUGH SETS ANALYSIS

To discover a rule-based model of the prediction from meteorological data, it was necesary to define the state vector at time instance *t* in terms of measured variables. The state vector had to be defined in such a way as to contain enough information to uniquely determine the next wind speed. The objective of the Rough Sets analysis was:

- to determine wheter the adopted definition of the state vector included sufficient and necessary information for derivation of the next value of the wind speed.
- to eliminate all redundant information (meteorological variables) from state vectors.
- to generate a set of rules describing the next value of wind speed.

3.1. Data Acquisition

The meteorological variables values are obtained from a meteorological station that has several channels. The station sends at fixed and customized intervals the information captured to a Personal Computer by means of RS-232 serial communication. In the P.C., installed as NT service, a program of data capture process the data coming from the serial port. The data are collected every 2 min.. These values are consolidated calculating the averages at intervals of 10 min.. The used variables are described in the following:

Channel	Meteorological Variable	Unit
1	Wind Speed	Km/h
2	Wind Gust	Km/h
3	Wind Direction	Deg. M
4	Air Temperature	Deg. C
5	Relative Humidity	%RH
6	Air Pressure	mb
7	Visibility	Km.
8	Sunshine Duration	Min.
9	Net Atmospheric Radiation	W/m^2
10	Rainfall	mm
11	Solar Radiation	W/m^2
12	Water Temperature	Deg. C

Fig. 1 Meterorological Variables

At the present time, we dispose of the data of 1999 (52622 records). A model was derived for prediction to 10 min. of wind speed. The objective of the model was predict wind speed in subsequent time instances based on current and past data of the 12 meteorological variables. The state vector was defined as a *time window* of length *n* containing information about current and past values of all meteorological channels. The time window length resulting was of 46 attributes.

3.2 Data Discretization.

One of premises of the rough sets theory is that reducing the precision of data representation reveals data regularities [7]. Consequently, the firts step in the analysis of process data was the discretization, i.e. replacing actual values with ranges of values based on the expert criteria.

Month	1	2	3	4	5	6	...	12
Quarter	≥1 ≤3	≥4 ≤6	≥7 ≤9	≥10 ≤12				
Hour of day	≥0 <8	≥8 <14	≥14 <8					
Wind Speed	≥0 ≤5	>5 ≤20	>20 ≤40	>40 ≤70	>70 ≤120	>120		
Wind Gust	≥0 ≤5	>5 ≤20	>20 ≤40	>40 ≤70	>70 ≤120	>120		
Wind Direction	>0 ≤90	>90 ≤180	>180 ≤270	>270 ≤360				
Air Temperature	>-20 ≤10	>10 ≤14,2	>14,2 ≤18,5	>18,5				
Relative Humidity	>0 ≤61,3	> 61,3 ≤72,4	>72,4 ≤83,4	>83,4				
Air Pressure	>800 ≤1007,9	>1007,9 ≤1016,6	>1016,6 ≤1025,1	>1025,1				
Visibility	≥≥≥0 ≤2,5	>2,5 ≤5	>5 ≤10	>10				
Sunshine Duration	0	>0 ≤5	>5					
Atmospheric Radiation	>-300 ≤-50	>-50 ≤40	>40 ≤340	>340				
Rain	0	>0 ≤0,41	>0,41					
Solar Radiation	>-300 ≤-15	>-15 ≤-5	>-5 ≤190	>190				
Water Temperature	>-20 ≤13,3	>13,3 ≤16	>16 ≤18,5	>18,5				

Fig. 2. Discretization of Meteorological Variables

3.1.1 Elimination of Redundant Variables

To eliminate redundant variables from the state vector, an attribute reduct was computed with respect to the wind speed.

The reducts almost did not eliminate attributes of the time window. In particular the calculation of the reduct eliminated only the following attributes:

- Visibility: two moments ago.
- Visibility: in the previous moment.
- Visibility: in the current moment.

We conclude that the Visibility is redundant.

3.1.2 Core

The calculated core has been:

- Hour of day: two moments ago.
- Rain: two moments ago.
- Hour of day: In the previous moment.
- Wind Speed: in the previous moment.
- Wind Direction: in the current moment.

4 ROUGH SETS AND DECISION RULE GENERATION

The Rough Sets theory is particularly well suited to deal with inconsistencies in the input decision table. If the input decision table is inconsistent then lower and upper approximations of the decision classes are computed. The decision rules are generated from these approximations. As a consequence of this way of treating inconsistencies two basic kinds of rules are distinguished (i.e. following the terminology used in [14], [16]): **Certain Rules** (also called exact, deterministic, or discriminant) and **Possible Rules** (also called approximate, non-deterministic)

For each decision class, certain decision rules are generated from objects belonging to its lower approximation. Possible decision rules are generated either from the upper approximation or from the boundaries of this decision class with other classes.

We have developed one extension of algorithm LEM3 based in the Stochastic Approximation Space to generate rules [15]. LEM3 is an algorithm for incremental learning of production rules from examples based in the boolean approximation space. In our approach [15] the lower and upper approximations are defined by using the concept of the Probabilistic (Stochastic) Approximate Classsification [18] with a level of certainty β ($0.5 \le \beta \le 1$).

A **Stochastic Approximation Space** is a triple StoAS = (*OB*, *R*, *P*), where *OB* is a universe, *R* is an equivalence relation on *OB*, and *P* is a probability measure on subsets of *OB*. In this approach, the lower and upper approximations of a subset *X* in *OB* are defined by using the concept of probabilistic approximation with a level of certainty β ($0.5 \le \beta \le 1$).

The ***R*-lower** and ***R*-upper β approximations** of *X* in *StoAS* are defined [17] as:
$\underline{R}X = \{ x \in U \mid P(X \mid [x]) \geq \beta \}$ and $\overline{R}X = \{ x \in U \mid \beta > P(X \mid [x]) \geq 0.5 \}$.

They belong to lower β–approx. of *X* those examples of which it is possible to be affirmed with a level of certainty β which they belong to class *X*.

In the Boolean Approximation Space an example is consistent o inconsistent. In StoAS, the consistency should be verified concerning a class and a certainty level β. An example could be consistent with a class and inconsistent with another, or it could be inconsistent with several classes.
It is said that an **example e_i is consistent** with a class *X* if the conditional probability $P(X|[e_i]) \geq \beta$. The conditional probability $P(X|[e_i])$ is defined as $P(X \cap [e_i])/P([e_i])$.

In the particular case of β=1, an example e_i is consistent with class *X*, if $P(X|[e_i]) = 1$; that is to say if $[e_i] \subseteq X$. This one is the condition of consistency in the BoolAS.
In the upper β-approx. definition we used the parameter α to instead of constant 0.5: $\overline{R}X = \{x \in U \mid P(X \mid [x]) \geq \alpha\}$, and we establish $\alpha = (1-\beta)$ to maintain the coherence with the meaning of certainty level β.

In [17] also considers a superior limit β in the precision of the upper β-approximation: $\overline{R}X = \{x \in U \mid \beta > P(X \mid [x]) \geq 0.5\}$. We did without this superior limit to conserve the coherence with the definition of upper approx. according to the theory of Rough Sets. The region that defines the previous expression is in fact the boundary region [7]. The generation of possible rules from the examples of the boundary region instead of the examples of the upper approximation has a disadvantage: rules are obtained that they only classify for the examples of the boundary region with which generalization is lost. And it has an advantage: between the rules generated from upper β-approx (considered as in [17]) will not be rules that already have been generated from the inferior approximation, with which we saved the later process of simplification of rules.

The method of rules generating is based on learning a **discriminant description**, i.e., to learn the smallest set of minimal rules, describing the concept. In LEM3 [17] and in our approach [15], rules are generated by using the non-incremental rules-generating algorithm LEM2 [16].

Let *B* be a nonempty lower or upper approximation of a concept represented by a decision-value pair (*d*,*w*). Set *B* **depends** a set *T* of attribute-value pairs *t* if and only if $\{ \emptyset \neq [T] = \bigcap_{t \in T} [t] \subseteq B \}$.

Set *T* is a **minimal complex** of *B* if and only if *B* depends on *T* and no proper subset *T'* of *T* such that B depends on *T'*. Let $\mathcal{T}$ be a nonempty collection of nonempty sets of attribute-value pairs. Then $\mathcal{T}$ is a **local covering** of *B* if and only if the following conditions are satisfied:

(1) each member *T* of $\mathcal{T}$ is a minimal complex of *B*,
(2) $\bigcup_{t \in T} [T] = B$, and
(3) $\mathcal{T}$ is minimal, i.e., $\mathcal{T}$ has the smallest possible number of members.

Procedure LEM2
(**input**: a set *B*,
output: a single local covering $\mathcal{T}$ of set *B*);
begin
$G := B$;
$\mathcal{T} := \emptyset$;
while $G \neq \emptyset$
begin
$T := \emptyset$;
$T(G) := \{ t \mid [t] \cap G \neq \emptyset \}$;
while $T = \emptyset$ **or** $[T] \not\subseteq B$
begin
select a pair pair $t \in T(G)$ with the highest attribute priority, if a tie occurs, select a pair $t \in T(G)$ such that $|[t] \cap T(G)|$ is maximum; if another tie occurs, select a pair $t \in T(G)$ with the smallest cardinality of $[t]$; if another tie occurs, select first pair;
$T := T \cup \{t\}$;
$G := [t] \cap G$;
$T(G) := \{ t \mid [t] \cap G \neq \emptyset \}$;

$T(G) := T(G) - T$;
end {while}
for each t in T **do**
if $[T - \{t\}] \subseteq B$ **then** $T := T - \{t\}$;
$\mathcal{T} := \mathcal{T} \cup \{T\}$;
$G := B - \cup_{T\in \mathcal{T}} [T]$;
end {while};
for each T in $\mathcal{T}$ **do**
if $\cup_{S\in \mathcal{T} -\{T\}}[S]=B$ **then** $\mathcal{T} := \mathcal{T} - \{T\}$;
end {procedure}.

In order to evaluate discovered rules several measures has been used:

Probability: conditional probability that the rule defines the concept (or class) that appears in the right part of the rule given the conditions in the left part of the same one. This probability is measured on the training set.
Specificity: the total number of attribute-value pairs on the left-hand side of the rule. The matching rules with a larger number of attribute-value pairs are considered more specific.
Strength: the total number of examples correctly classified by the rule during training.
TrMatch: the total number of training examples that match with the left-hand side of the rule.

A subset of the generated rules (923) from data of year 1999 is shown in Appendix A. The quality of approximation has been 96% and the accuracy of classification has been 100%.

5 EXPERIMENT WITH A TEST SET: DATA OF FIRST QUARTER OF YEAR 2000

Typically, a rule set induced from training data is used for classsifications of unseen (sometimes called testing) cases. In [16] two different classification schemes are used. The first one will be called **Naive**. If no rule exists that matches a new, unseen example or if two (or more) rules match the example but these rules indicate different concepts, the example is treated as an error. The second one is called **New**, and it is a modification of the "bucket brigade algorithm" [19,20]. In New approach, the decision to which concept an example belongs is made using three factors: *Strenght*, *Specificity*, and *Support*.

In this work both systems of classification have been implemented, but with certain modifications that mainly improve the results in the Naive method. In the previous section already it has been commented that the induced rules come accompanied from four measures: Probability, Specificity, Strength and TrMatch. The first measurement is used in the Naive method for the selection of the rule associated for the new example to classify. The other three measures are used in the New method to calculate a factor that will determine the class of the new example. The application of the improvement of the Naive method consists of the following steps:

For each example to classify, all the rules are computed that match with the example. Two situations can appear:

- *All the rules that match with the example indicate the same concept or class.* In this case the example is classified like pertaining to the concept indicated by the rules.
- *The rules that match with the example indicate different concepts.* The example is classified like pertaining to the concept of that rule that has greater probability. In the particular case that there are several rules with the same probability -greater than anyone of the rest of rules- and that indicate different concepts, the concept is chosen that is associated to the rule of greater number of pairs attribute-value. If the attribute-value pairs number of the rules that indicate different concepts with equal probability is the same one, is chosen the first concept.

If an example does not match completely with some rule, a concept by defect is assigned to him. This is equivalent to the application of the rule by defect, is to say a rule without antecedents that a certain class or concept concludes.

In New method the decision to choose the concept to which an example belongs does on the basis of the factor *Support*, that defines by means of the factors *Strenght* and *Specificity*:

$$Support = \sum_{\text{matching rules } R \text{ describing } C} Strength(R) * Specificity(R)$$

Support is the sum of the votes of those rules that match with the example and indicate the same concept.

In New method the classification of a new example is not done choosing the best rule than match with the example, but that the decision becomes according to the global knowledge that stores the set of rules that match with the example.

In Naive method when no rule match completely with the example the rule by defect is applied.

In New classification, if matching complete is not possible, all the rules that match partially (rules that match with at least a pair a-v of the example) are identified.

For each rule R that match partially is calculated the Matching factor, $Tr\_Match(R)$. If *Support* is null for all the concepts, then the following expression for each concept is calculated:

$$\sum_{\text{partially matching rules } R \text{ describing } C} Tr\_Match(R) * Strength(R) * Specificity(R)$$

Like in the previous case, the concept C that is assigned to the new example to classify is the one that has the greater value for the previous expression.

With this system of classification it is not necessary to generate a class by defect in the moment of the rules induction .

The information about the set of rules generated is shown next:

***Training Set:* Year 1999**

Number of training examples: 44220 (it is necessary to consider that the data acquisition computer had some shutdown during year 1999)

α: 0.5 (the inconsistent examples are including within the upper approximation of that class -if there is some- to belong at least half of the examples)

β: 0.5 (the inconsistent examples whose consistency with a class is superior to 0.5 are considered within their lower approximation)

Noise Level R: 0.9 (a level of noise of 90% has been introduced to avoid *overfitting*. Thus it is allowed that each generated rule classifies of correct form a certain percentage *R* (90%) of training examples and consequently classifies of erronea form a percentage *1-R* (10%) of training examples.)

Unknown Values: yes

Attributes Number: 45

Classes Number: 6

Approximation Quality: 96%

Approximation Accuracy: 100%

Number of Rules: 923

Fig. 3. Information of the set of rules

The classification results of applying the set of rules on different testing sets are shown next.

	Number of testing examples	Percentage of successes (Naive classification)	Percentage of successes (New classification)
1^{th} trimester 1999	10930	76,5%	76,7%
2^{th} trimester 1999	10438	74,7%	72,9%
3^{th} trimester 1999	9982	75,6%	76 ,4%
4^{th} trimester 1999	12870	76,9%	71,1%
1^{th} trimester 2000	11493	76,7%	69,2%

Fig. 4. Results of applying the set of rules on different testing sets

6 CONCLUSIONS

We have introduced a representation for dealing with time series. The transformation from Time Information System to Information System is significant since it enables us to use standard Rough Sets techniques for further processing of the data.

As a result of the test that it has been made, it has been arrived to the conclusion that in order to carry out a prediction with enough quality it is essential using the values of the following attributes:

- Hour of day: two moments ago.
- Rain: two moments ago.
- Hour of day: in the previous moment.
- Wind Speed: in the previous moment.
- Wind Direction: in the current moment.

And was observed that visibility channel has no influence in wind speed prediction.

The **approximation quality** has been of 96%. That is to say, 96% of all training examples can be correctly classified in classes using the knowledge contained in the complet attributes set. The **approximation accuracy** has been of 100%. That is to say, there is the 100% of posible correct decisions when classifying examples employing the knowledge contained in the complet attributes set.

A **set of rules** (923) from a **training set** of 44220 records corresponding to year 1999 has been generated. The training records contain the values of 12 meteorological variables registered every 10 minutes.

The **percentage of successes** on **testing set** (1^{th} trimester of year 2000) has been of **76,7%** with method Naive of classification. It is to say, three of each four predictions are made successfully on the testing set.

These calculations are based on the exposition to eliminate of the superior approach of a class those examples whose conditional probability to the class is minor than 0.5 (*certainty level* β=0.5) and on a *noise level* of 0.9.

REFERENCES

[1] J.G. Bazan, A. Skowron and P. Synak, Market Data Analysis: A Rough Set Approach, Technical Report from the University of Warsaw, (1994).

[2] A. Bjorvand and J. Komorowski, 'Practical Applications of Genetic Algorithms for Eficient Reduct Computation', 15th IMACS World Congress 1997 on Scientific Computation, Modelling and Applied Mathematics, (1997).

[3] R. Golan and D. Edwards, *Temporal Rules Discovery using Datalogic/R+ with Stock Market Data*, Rough

Sets, Fuzzy Sets and Knowledge Discovery, Edited by W. Ziarko (ed.), 74-81, Springer-Verlag, (1993).

[4] R. Golan and W. Ziarko, *A Methodology for Stock Market Analysis Using Rough Set Theory*, Proceedings of IEEE/IAFE Conference on Computational Intelligence for Financial Engineering, New York City, 32-40, (1995).

[5] G. Janacek and L. Swift, *Time Series -Forecasting, Simulation, Applications*, Ellis Horwood, 1993.

[6] J.S. Ostroff, *Temporal Logic for Real-Time Systems Research Studies*, Press LTD, John Wiley & Sons Inc., (1989).

[7] Z. Pawlak, *Rough Sets - Theoretical Aspects of Reasoning About Data*, Kluwer Academic Publishers, Dordrecht, (1991).

[8] A. Skowron, *Synthesis of Adaptive Decision Systems from Experimental Data*, SCAI'95 - Proceedings of the Fifth Scandinavian Conference on Artificial Intelligence, A. Aamodt and J. Komorowski (eds.), IOS Press, (1995).

[9] W. Zakowski, Sequences of Information Systems, Configurations and Conflicts, *Bull. Pol. Acad. Sci. Ser., Tech.*, **41**, 295-304, (1993).

[10] W. Ziarko, R. Golan and D. Edwards, An application of Datalogic/R Knowledge Discovery Tool to Identify Strong Predictive Rules In Stock Market Data, AAAI-93 Workshop on Knowledge Discovery in Databases, (1993).

[11] C. Pérez-Llera, M.C. Fernández-Baizán, E. Menasalvas, J. Feito, M. Luengo, *Prediction of Meteorological Variables*, Proc. 4th Intern. Conf. On Information Systems, Analysis and Synthesis, 281-284. N. Callaos, T. Yang, J. Aguilar (eds.), International Institute of Informatics, Orlando (USA), (1998).

[12] C. Pérez-Llera, M.C. Fernández-Baizán, J. Feito, M. Luengo, Local Short-Term Prediction of Wind Speed: A Neural Network Analysis, submitted to *Environmental Modelling & Software, Elsevier*.

[13] Z. Pawlak, R. Slowinski, Rough Sets Approach to Multi-attribute Decision Analyis, *European Journal of Operational Research,* **72**, 443-459, (1994).

[14] W. Ziarko, Review of basis of rough sets in the context of Data Mining, Proc. IV Intern. Workshop on Rough Sets, Fuzzy Sets and Machine Discovery, S. Tsumoto, S. Kobayashi, T. Yokomo, H. Tanaka (eds.), Tokyo, 447-457 (1996).

[15] M.C. Fernández-Baizán, C. Pérez-Llera, J. Feito, J., A. Almeida, LEM3 algorithm generalization based on Stochastic Approximation Space, submitted to RSCTC 2000.

[16] J.W. Grzymala-Busse, A New Version of the Rule Induction System LERS, *Fundamenta Informaticae*, **31**, 27-39, (1997).

[17] Chan, C.C.: Incremental Learning of Production Rules from Examples under Uncertainty: A Rough Set Approach. Int. J. of Software Engineering and Knowledge Engineering **1** 4 (1991) 439-461

[18] Wong, S.K.M., Ziarko, W.: INFER-An Adaptative Decision Support System Based on the Probabilistic Approximate Classsification. In Proc. 6th Int. Workshop on Expert Systems and their Applications (1986) 713-7262

[19] Gunn, J.D., Grzymala-Busse, .W.: Global tempearture stability by rule induction: an interdisciplinary bridge. Human Ecology 22 (1994) 59-81

[20] Holland, J.H., Holyoak, R.E.: Induction, Processes of Inference, Learning, and Discovery. MIT Press, Cambridge MA (1986)

Appendix A

CERTAIN RULES FOR CLASS
Wind Speed between 0 .. 5 Km/h

IF (Month 20’ ago, January) &
(Wind Direction 20’ ago, 180 .. 270 Deg. M) &
(Hour of Day at present time, Evening) &
(Wind Gust at present time, 0 .. 5 Km/h)
THEN (Wind Speed within 10’, 0 .. 5 Km/h)
Probability: 1.000000
Specificity: 4
Strength: 467
TrMatch: 467

IF (Month 20’ ago, January) &
(Hour of Day 20’ ago, Morning) &
(Wind Gust at present time, 0 .. 5 Km/h)
THEN (Wind Speed within 10’, 0 .. 5 Km/h)
Probability: 1.000000
Specificity: 3
Strength: 469
TrMatch: 469

IF (Month 20’ ago, January) &
(Hour of Day 20’ ago, Evening) &
(Pressure 20’ ago, 1007,9 .. 1016,6 mb)
THEN (Wind Speed within 10’, 0 .. 5 Km/h)
Probability: 0.854685
Specificity: 3
Strength: 447
TrMatch: 523

CERTAIN RULES FOR CLASS
Wind Speed between 5 – 20 Km/h:

IF (Quarter 20’ ago, 2th Trimester) &
(Air Temperature 20’ ago, 14,2 .. 18,5 Deg. C) &
(Humidity 20’ ago, 61,3 .. 72,4 %RH) &
(Air Temperature 10’ ago, 14,2 ..18,5 Deg. C) &
(Water Temperature 10’ ago, 16 .. 18,5 Deg. C) &
(Wind Direction at present time, 0 .. 90 Deg. M) &
(Air Temperature at present time, 14,2 .. 18,5 Deg. C)
THEN (Wind Speed within 10’, 5 .. 20 Km/h)
Probability: 0.936275
Specificity: 7
Strength: 191
TrMatch: 204

IF (Month 20’ ago, December) &
(Our of Day 20’ ago, Evening) &
(Wind Speed 20’ ago, 5 .. 20 Km/h) &

(Humidity 20’ ago, 61,3 ..72,4 %RH) &
(Wind Speed at present time, 5 .. 20 Km/h) &
(Pressure at present time, 1016,6 .. 1025,1 mb)
THEN (Wind Speed within 10’, 5 .. 20 Km/h)
Probability: 0.992481
Specificity: 6
Strength: 264
TrMatch: 266

IF (Month 10’ ago, September) &
(Wind Speed at present time, 5 .. 20 Km/h) &
(Wind Direction at present time, 180 .. 270 Deg. M) &
(Pressure at present time, <1007,9 mb) &
THEN (Wind Speed within 10’, 5 .. 20 Km/h)
Probability: 0.884615
Specificity: 4
Strength: 207
TrMatch: 234

CERTAIN RULES FOR CLASS
Wind Speed between 20 –40 Km/h

IF (Wind Direction 20’ ago, 180 .. 270 Deg. M) &
(Pressure 20’ ago, 1016,6 .. 1025,1 mb) &
(Wind Gust 10’ ago, 40 .. 70 Km/h) &
(Wind Direction 10’ ago, 180 .. 270 Deg. M) &
(Wind Direction at present time, 180 .. 270 Deg. M) &
(Humidity at present time, <61,3 %RH)
THEN (Wind Speed within 10’, 20 .. 40 Km/h)
Probability: 1.000000
Specificity: 6
Strength: 10
TrMatch: 10

IF (Air Temperature 20’ ago, >18,5 Deg. C) &
(Atmospheric Radiation 10’ ago, -300 .. -50 W/m^2) &
(Water Temperature10’ ago, 13,3 .. 16 Deg. C) &
(Wind Speed at present time, 20 .. 40 Km/h) &
(Solar Radiation at present time, 0 .. 5 W/m^2)
THEN (Wind Speed within 10’, 20 .. 40 Km/h)
Probability: 1.000000
Specificity: 5
Strength: 12
TrMatch: 12

IF (Hour of Day 20’ ago, Morning) &
(Air Temperature 20’ ago, 10 .. 14,2 Deg. C) &
(Solar Radiation 20’ ago, -15 .. -5 W/m^2) &
(Solar Radiation 10’ ago, -15 .. -5 W/m^2) &
(Month at present time, December) &
(Wind Speed at present time, 20 .. 40 Km/h) &
(Solar Radiation at present time, -15 .. -5 W/m^2)
THEN (Wind Speed within 10’, 20 .. 40 Km/h)
Probability: 1.000000
Specificity: 7
Strength: 18
TrMatch: 18

CERTAIN RULES FOR CLASS
Wind Speed between 40-70 km/h

IF (Hour of Day 20’ ago, Morning) &
(Rain 20’ ago, >0,41 mm.) &
(Atmospheric Radiation 10’ ago, -50 .. 40 W/m^2) &
(Wind Speed at present time, 40 .. 70 Km/h)
THEN (Wind Speed within 10’,40 .. 70 Km/h)
Probability: 1.000000
Specificity: 4
Strength: 6
TrMatch: 6

IF (Solar Radiation 20’ ago, -15 .. -5 W/m^2) &
(Wind Speed 10’ ago, 40 .. 70 Km/h) &
(Water Temperature 10’ ago, <=13,3 Deg. C) &
(Wind Speed at present time, 40 .. 70 Km/h) &
(Wind Gust at present time, 70 .. 120 Km/h)
THEN (Wind Speed within 10’, 40 .. 70 Km/h)
Probability: 1.000000
Specificity: 5
Strength: 10
TrMatch: 10

IF (Wind Speed 20’ ago, 20 .. 40 Km/h,) &
(Wind Direction 10’ ago, 270 .. 360 Deg. M) &
(Water Temperature 10’ ago, <=13,3 Deg. C) &
(Wind Speed at present time, 40 .. 70 Km/h) &
(Solar Radiation at present time, -15 .. -5 W/m^2)
THEN (Wind Speed within 10’, 40 .. 70 Km/h)
Probability: 1.000000
Specificity: 5
Strength: 6
TrMatch: 6

CERTAIN RULES FOR CLASS
***Wind Speed* between 70-120 Km/h**

IF (Wind Speed 20’ ago, 40 .. 70 Km/h) &
(Wind Speed 10’ ago, 40..70 Km/h) &
(Wind Speed at the present time, 40 .. 70 Km/h) &
(Wind Gust at the present time, >120 km/h)
THEN (Wind Speed within 10’, 70 .. 120 km/h)
Probability: 1.000000
Specificity: 4
Strength: 1
TrMatch: 1

IF (Wind Gust 20’ ago, >120 Km/h) &
(Wind Speed 10’ ago, 70 .. 120 Km/h) &
(Wind Gust 10’ ago, 70 .. 120 Km/h)
THEN (Wind Speed within 10’, 70 .. 120 Km/h)
Probability: 1.000000
Specificity: 3
Strength: 1
TrMatch: 1

IF (Wind Direction 20’ ago, 270 .. 360 Km/h) &
(Sunshine Durat. 20’ ago, >5 Min) &
(Solar Radiation 20’ ago, >190 W/m^2) &
(Wind Gust 10’ ago, 70 .. 120 Km/h) &
(Solar Radiation at present time, -5 .. 190 W/m^2)
THEN (Wind Speed within 10’, 70 .. 120 Km/h)
Probability: 1.000000
Specificity: 5
Strength: 1
TrMatch: 1
