Knowledge of Administration to support the taking of Decisions to the Contamination in the city of Santiago from Chile

Santiago Zapata Cáceres, Luis Escobar Ramirez
Department of Computer Science
Engineering Faculty
Metropolitan Technological University of Chile
szapata@utem.cl, laescoba@utem.cl

Abstract: - The contamination of the air is one of the biggest problems that affects the countries in almost any part of the world. The increment in the quantities of gases and particles potentially harmful for the health and the environment has been verified to world scale, and every day is but I obviate that the answer to these problems concentrates in the administration of knowledge to support the taking of decisions.

In Chile, for law, the obligation settles down of developing descontaminación plans. During the period autumn-winter the population of Santiago's city, it is affected by a sudden increase in the levels of contamination of the air. This situation is known as critical episode of atmospheric contamination. These episodes they originate starting from the convergence of a series of meteorological factors that you/they impede the good ventilation of Santiago's y/o basin due to an increment in the emissions of previous pollutants to the event.

The approach to decree a critical episode of contamination this referred to the index ICAP that is generated by the data of the Net of Monitoreo of gases and particles that are entered a Model of Presage

This investigation looks for to find intelligent knowledge for the prediction and control of environmental problems in the city of Santiago from Chile.

Key-Words: - Knowledge of Administration, Decisions Support System (DSS), Knowledge Data Discovery (KDD), Contamination.

1 Introduction

The contamination of the air in Santiago's city originates mainly during the period autumn-winter, in these months the population of the Metropolitan Region is affected by a sudden increase in the levels of contamination of the air.

These situations, known generically as critical episodes of atmospheric contamination, take place when they register high levels of certain polluting agents' concentration during periods of short duration. These episodes originate starting from the convergence of a series of meteorological factors that you/they impede the good ventilation of Santiago's basin and also due to an increment in the emissions of the polluting agents.

The Index of Quality of the Air for Material Particulado, ICAP [1], [2], [3] it is the indicator that through the emanated data of the presage pattern Cassmassi serves as antecedent so that government's authority can determine yes Santiago's city this in presence of a critical episode of contamination. The pattern Cassmassi predicts the maximum value of concentration average of 24 hours of material breathable particulado PM10, for the period of 00 at 24 hours of the following day.

- For this presage of episodes a team of expert meteorologists prepares the presage of meteorological conditions associated to episodes of atmospheric contamination daily for the Metropolitan Region, based on modernized meteorological information, national and international, and to the data of quality of air of the net of automatic monitoreo of quality of the air and meteorology Net MACAM [2], [45].

- It is important to highlight that because the pattern predicts the quality of the air of the following day, the declaration of an episode on the part of the authority it doesn't imply that the air has worsened, but rather it could end up worsening. That is to say, the episodes are decreed in preventive form to avoid to reach the predicted indexes, and this way to protect the population's health.

This form of to approach the problem and to make decisions presents some undesirable characteristics, such as;
- The pattern of Cassmassi considers eight monitoreo stations in isolated form and in many occasions it has been observed that during the occurrence of critical episodes, the stations interact to each other.

- The permanency of the point of view reduccionista (in I deteriorate of the systemic focus) it has induced the occurrence of countless prediction errors.

- In a study carried out by the CONAMA [3], [48], [49] he/she showed that this model is ineffectivo to predict critical levels of contamination. For one period understood among March 20 from 2003 to August 10 2003 it has only guessed right 8 times in the prediction of critical situations (it alerts and preemergencia), with 30 false alarms and 12 underestimated episodes.

- Due to the global climatic change, every year is more difficult to carry out reliable presage, but this model has not adapted for these new situations (that is to say, you/he/she maintains her stable parameters, although the conditions in study change).

- Another element that stands out is that this model of Cassmassi has not been modernized since you/he/she was designed, and they have already happened but of six years from her creation, is also ignored that the technicians of the Cenma, organism of the University of Chile in charge of carrying out the presage for the taking of decisions, they are requesting her renovation for some time.

- With regard to the primary norm of quality of the air for material thick particulado (PM10) it is said that the current pattern of prediction only centers the attention in the values means and not in the variations hour per hour, what means that the maxima of contamination, in every day, are hidden in the average of 24 hours.

2 Problem Formulation

At the present time and to world level, the study and control of problems of environmental contamination, it is approached with the support of the calls DSS (Decision Support Systems, Systems of Support of Decisions in Castilian) that are a specific class of systems of information that support the processes of taking of decisions in the organizations. These DSS is interactive systems that help to the taking of decisions facilitating the handling of the data, documents, knowledge model that are used to solve problems inside the organizations.

One of the environments in that the DSS is of great utility is in the support to the taking of decisions for analysis and control of environmental problems, in particular the support that can offer the Mining of Data (Mining Data) when allowing to extract patterns, models, relationships, tendencies, etc. that finally allow to find " rules " or " patterns " (" knowledge ") starting from the data and to communicate them to the user through the DSS.

The general objective of our investigation is to propose a methodology based at the moment on DSS like an alternative to the method used to measure and to make decisions with regard to the environmental emergencies in Santiago's city.

2.1 CURRENT ADMINISTRATION OF THE CONTAMINATION IN THE CITY DE SANTIAGO

One of the big problems that possesses the Metropolitan Region, especially the commune of Santiago from Chile is the concentration in the atmosphere of polluting particles that you/they damage the health of people that live and they traffic for this city. For this reason it is important to have a good support system to the decisions that it allows to analyze the data to obtain results that they help to make better decisions of prevention for the citizenship.

A responsible entity that takes charge of looking after the care of the environment is the SESMA [44] who through a net of mensuration called Net MACAM [45] they go registering the levels of contamination daily in the environment.

2.2. MACAM Net

The official net of Automatic Monitoreo of Quality of the Air and Meteorology of the city of Santiago from Chile (Net MACAM) it was started in 1988 with 5 monitoreo stations, for the most part located in Santiago's central sector. In the winter of 1997, this net was renovated and enlarged to 8 stations of automatic monitoreo, in what today is denominated as Net MACAM-2, clerk of the Metropolitan Service of Health of the Atmosphere (SESMA). The place of Internet that allows to obtain the data is the following one:

http://www.minsal.cl/sesma/pvcasesma/default.htm
2.3. Chilean norms of Quality of the Air

2.3.1. Norms

The Resolution Nº369, of the Ministry of Health of April of 1988, published in the official newspaper April of 1988, establishes the indexes of quality of the air to determine the level of atmospheric contamination of the Metropolitan Region.

2.3.2 ICA E ICAP Index

As much the ICA as the ICAP, they give origin to the following qualification of the situation according to the obtained value:

<table>
<thead>
<tr>
<th>INDEX</th>
<th>QUALIFICATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 – 100</td>
<td>GOOD</td>
</tr>
<tr>
<td>101 – 200</td>
<td>REGULATE</td>
</tr>
<tr>
<td>201 – 300</td>
<td>BAD</td>
</tr>
<tr>
<td>301 – 400</td>
<td>CRITICIZE</td>
</tr>
<tr>
<td>401 – 500</td>
<td>DANGEROUS</td>
</tr>
</tbody>
</table>

2.4. Model Predictivo of Joseph Cassmassi

2.4.1. Introduction

This model is used to predict the episodes of Alert, Preemergercia or Emergency whose measures that she takes the authority have been described in the introduction of the present investigation, the information surrended by the pattern implies to be ahead to the excessive increase of the Indexes of Quality of the Air considering the meteorological variables and the levels of contamination for Material Particulado in suspension.

The Pattern of Prediction uses eight monitor stations distributed geographically in the city of Santiago from Chile and it stops its estimate it considers: the historical behavior of the atmospheric contamination and their association with meteorological parameters, levels of atmospheric contamination and their daily variation. All these factors are entered to a mathematical Polynomial that was developed by the American physique Joseph Cassmassi in the year 1999, based on classic statistical techniques of multiple lineal regression [47].

2.4.2. I model of presage for episodes

The Plan of Prevention and Atmospheric Descontaminación for the Metropolitan Region (PPDA) it established the necessity to create a model of presage of episodes of contamination, due to the importance of advancing the control measures and of receipt to the population in front of the occurrence of critical situations.

On the other hand, the Supreme Ordinance N° 59 of 1998 of the Ministry would Secrete General of the Presidency that establishes the norm of primary quality for material breathable particulado (MP10), it included the requirements that it should satisfy a methodology of presage of quality of air.

In accordance with this requirements, by means of resolution Nº12.612 of 1998 of the Metropolitan Service of Health of the Atmosphere of the Metropolitan Region (SESMA), you oficializo the first application of a model of presage of quality of air in Santiago's city, in use starting from July of 1998.

As a form of optimizing this tool, in 1999 CONAMA he/she took charge a study to improve the methodology of presage of quality of air in the Metropolitan Region. The study gave a new presage model, denominated model Cassmassi as a result.

2.4.3. Used variables

The methodology of presage of concentrations of PM10 is based on calculation algorithms developed by means of application of statistical techniques of multiple regression, focused to find relationships between possible variable predictivas and the variable to predict. The predictores includes observed meteorological variables, indexes of observed meteorological conditions and predicted, observed concentrations of pollutants, indexes of prospective variations of emissions and others.

2.4.4. Meteorological potential of atmospheric contamination (used Algorithms)

The operational application of this methodology considers two prediction algorithms for each station monitora:
A first algorithm includes the Index of Meteorological Potential of Atmospheric Contamination" (PMCA)
predicted for the following day. The figure 6.1 sample that the PMCA has a good correspondence with the averages of 24 hours of the observed concentrations of PM10.

Figures 2.1. - Relationship PMCA and contamination for PM10.

The second algorithm is only based on observations (of the same day and of the previous day). This way, if the first algorithm you cannot apply for inadequacy of the information, the second algorithm is used.

2.4.5. Validation of the pattern

In 1999/2000, was carried out an evaluation of the pattern Cassmassi on the part of the Department of Geophysics of the University of Chile. For it, the information of the period was used that extends from April 1 to September 17, 1999. This way, the independence of the data was guaranteed used to validate the pattern of those used in its construction. The validation shows that the pattern Cassmassi fulfills the requirements of the Supreme Ordinance No. 59 of 1998.

The detailed description of the pattern Cassmassi is in the study Improvement of the Forecast of Air Quality and of the Knowledge of Local the Meteorological Conditions in the Metropolitan Region" (Cassmassi 1999), [48], [49].

2.5. Flaws of the Pattern Predictivo of Cassmassi

The pattern predictivo is inefficient to predict critical levels of contamination, for example, in the year 2003 alone it guessed right 8 times the prediction of situations you criticize, with 30 false alarms and 12 underestimated episodes. According to studies (it Has That it Happens, May 16 the 2003), [50] it indicates that one of the main causes of the flaws is that it has not been modernized in 5 years, and that besides using the pattern predictivo of Cassmassi officially, leans on in other five non official models, to estimate the degree of certainty of the first one. For example, one day anyone the presage indicated that there was 80% of probabilities of alert environmental, but with an incertitude of 35% that he became environmental preemergencia.

This model considers 8 monitoreo stations in isolated form and in many occasions has been observed that during the occurrence of critical episodes, the stations interactúan to each other. Adopting the focus reduccionista instead of the systemic focus has generated the occurrence of countless errors.

2.6. Decisions starting from the results of the pattern

Different they are the noxious effects that he/she can bring to the health of people the contamination of the air that at the moment faces Santiago's city. However, these you specially dangerous tornan in the case of exhibitions at very high levels during short periods of duration, and that it is what happens when we are in front of a critical episode. It is then fundamental to be able to be early to the situations of more risk to stop the negative consequences that it carries an environmental emergency.

The results of the pattern presage are used as tool of preventive information on the part of the regional authorities, this way, if the value hurtled by the pattern presage overcomes in some of the eight stations:

- The level ICAP 200 according to this model should be decreed it Alerts Environmental
- The level ICAP 300 according to this model should be decreed it Environmental Preemergencia
- The level ICAP 500 according to this model should be decreed it Environmental Emergency

2.7. Cost of the errors of the Pattern Predictivo

According to economic studies, the costs associated to the errors of the pattern predictivo overcame those $2.000 million pesos in the year 2001 about 2.7 million Euros, on those $5.000 million pesos in the year 2002 about 6.9 million Euros, and was considered that for the year 2003 this figure could be duplicated reaching near values to the 12 million Dollars, [50], [52].

As having measured correctivas, these studies recommended to revise the analysis of effectiveness of the measures adopted during the critical episodes; to evaluate the costs associated to the paralización of
industrial sources and in third place to revise and to improve the quality predictiva of the used models.

3. PROPOSAL OF WORK

3.1. Introduction

According to what shown previously for the administration of the problems caused by the environmental contamination in Santiago's city, we conclude that to the problem a solution can be contributed like we detail next.

The current treatment of the presented problem you can schematize as it shows it the figure 3.1., in which can stand out that the pattern of Cassmassi (it uses classic statistical tools), it has not been modernized approximately six years ago from their initial conception, it doesn't justify the critical episodes clearly and it bases their prediction on an average of behavior of the involved variables of the previous day, it is not systemic since it doesn't relate the effect sinergético of the data generated by the different stations of registration of data that are distributed geographically in Santiago's city.

Fig. 3.1. Current treatment of administration of environmental decisions

The alternative proposal, intends to prepare the basic data surrendered by the stations of monitoreo of data of the Net MACAM and the meteorological information of the prediction of the climate surrendered by the meteorological services of the city, in a DataWarehouse based on a model it Starchema.

The DataWarehouse should be designed to respond to consultations (not foreseen) relative to the activities of the organization analyzed from different points of view.

For example:

- I number of preemergencias decreed in winter.
- You alert decreed assertively.
- Commune with but preemergencias during the year, contained by station of the year

3.2. Design of the DataWarehouse SESMA

For each activity of the organization to be analyzed should design an outline it shatters or flake of snow, (Starchema and Snowflake). In this case the activity to analyze is the SITUATION, then this it will be our chart of facts which is the one that contains the important information of our organization.

![Fig. 3.2. Chart of having Made of the Pattern Shatters](image)

This chart contains the columns id_situación, id_medicion and id_comuna, it also contains the columns cost that it indicates us the cost associated to a certain situation, the column benefit that indicates us the benefit associated to a certain situation more the columns date and mensuration that it indicates us the mensuration made in a certain day (this mensuration was calculated previously of the original database OLTP).

Each line of the chart of facts contains the data observables (cost, benefit, mensuration) of the activity and of the references to the dimensions that characterize them (you nail other people's: id_medicion, id:comuna, dates).

A chart of facts generally represents a relationship many it is many with its respective dimensions.

The second step is to define our charts dimensions and its properties, they are the excellent dimensions for our consultations.

The dimensions that we define for our DataWarehouse are: mensuration, commune and a dimension of time dates.

With this and after desnormalizar (desnormalizamos to avoid the concatenation of charts when making consultations, this lends us a simpler use but with redundancy reason why is in charge of a lot of space) the dimensions, our outline shatters it is like it is shown in the figure 3.3:
3.3. YOU CONSULT TO THE DATAWAREHOUSE

We are under conditions of making consultations, for example, to show date in that the "situation" is "preemergencia" in the commune of "Pudahuel" in the "period" of "winter", this would be made in the following way:

```sql
select medicion.nombre, fecha.fecha from situacion, medicion, comuna, fecha
where situacion.id_medicion=medicion.id_medicion
and medicion.nombre='preemergencia' and
situacion.id_comuna=comuna.id_comuna
and fecha.id_fecha=situacion.fecha
and fecha.estacion='tinvierno'
```

Fig. 3.4. Result when consulting date and situation in period of winter in the commune of Pudahuel.

In the previous case, the dimension of time was introduced it dates in the analysis and we include a new aggregation level on the date type, this is known as drill-down.

Another consultation that being of interest could us to generate reports that allow to take serious decisions: I number of preemergencias decreed in winter, in code this consultation would be expressed in the following way:

```sql
select COUNT(*)
From situacion, medicion, comuna, fecha
where situacion.id_medicion=medicion.id_medicion
and situacion.id_comuna=comuna.id_comuna
and fecha.id_fecha=situacion.fecha
and medicion.nombre='preemergencia'
and fecha.estacion='tinvierno'
```

Fig. 3.5. Result when consulting preemergencias number decreed in winter.

In relation to the previous consultation we could show the communes and the date in that preemergencia was decreed during the winter. In code this consultation would be expressed in the following way:

```sql
select fecha.fecha, comuna.nombre
From situacion, medicion, comuna, fecha
where situacion.id_medicion=medicion.id_medicion
and situacion.id_comuna=comuna.id_comuna
and fecha.id_fecha=situacion.fecha
and medicion.nombre='preemergencia'
and fecha.estacion='tinvierno'
```

Fig. 3.6. Been by screen it consults it dates and commune with preemergencia decreed in winter.

This design of DataWarehouse allows us to know important data for the taking of Decisions Preposition the quality of the air in Santiago from Chile, decisions based on reports, graphics, etc.

The thick of the investigation besides the design of the DataWarehouse this in design of a powerful system to negotiate the data and that can give useful and reliable information for the taking of decisions.
3.4. CONCLUSIONS

The consultations on a DataWarehouse are a lot but simple that in a database relacional since a lot of information you filter before filling the chart of facts and their respective dimensions also taking into account the great help that it provides us the desnormalización to obtain a code a lot but simple and that gives us data that constitute excellent information for the taking of decisions.

The future work is promissory, because to find alternative solutions to the problem of the contamination in the city of Santiago from Chile has become a crusade of humanitarian character, since the problems to the population's health and the economy of the country are reaching levels that surpass the acceptable thing for the community.

5. REFERENCES

15] Diana Cristina Romero Sánchez, La Importancia De Los GDSS En El Trabajo Colaborativo, dromero@campus.zac.itesm.mx http://www.netmedia.info/netmedia/articulos.php , Netmedia S.A.


28] Artículo de Internet con el título “Data Mining: torturando los datos hasta que confiesen”, por Luis Carlos Molina Félix, coordinador del programa de data Mining (UCO), 2001


43] http://www2.ing.puc.cl/gescopp/investigacion.html

44] SESMA (Servicio de Salud Metropolitano del Ambiente), www.sesma.cl

45] RED MACAM (Red de Monitoreo Automático de Calidad del Aire y Meteorología), www.conama.cl/rm/568/article-1114.html


50] http://www.quepasa.cl/revista/2003/05/16/t-16.05.QP.NAC.CONTAMINACION.html


54] WEKA (Waikato Environment for Knowledge Analysis) http://www.cs.waikato.ac.nz/~ml/