An Intelligent System for Hydro Power Network Operations Planning *

Hakim Lounis
Department of Computer Science
Université du Québec à Montréal,
Canada
lounis.hakim@uqam.ca

Kaddour Boukerche
Centre de Recherche Informatique
de Montréal (CRIM),
Canada
kaddour.boukerche@crim.ca

Houari A. Sahraoui
Department of Computer Science and
Operations Research
Université de Montréal, Canada
sahraouh@iro.umontreal.ca

Abstract - **This paper presents a reengineering work tending to increase to a significant degree some software qualities relevant in the management of production of a hydroelectric network. An object-oriented knowledge-based architecture is proposed to ensure an intelligent and automatic management of the knowledge in use in the daily decisional process of a major Canadian company. It includes a machine-learning module for historical data processing.**

I. INTRODUCTION

Knowledge-Based Systems (KBS) are used in numerous application domains, one of which is the field of hydroelectricity, in which this work fits [1] [2] [3]. KBS are used to reproduce an expert's reasoning and are based on two distinct components: knowledge and reasoning. Separation between these two levels of intervention makes it possible to offer a flexibility of operation that many traditional software approaches are missing. KBS are presently an effective and useful solution to integrate the necessary analyses of hydropower experts and to meet the needs of the hydroelectric industry.

An essential requirement of the KBS design process is the use of efficient representations of large amounts of knowledge; this ensures the consistency and effective exploitation of the KBS algorithms. The available knowledge can be explicit or implicit. An explicit representation consists of a symbolic expression of human expert knowledge; an implicit representation is knowledge that is usually hidden in data. It requires further processing of the data before useful information can be extracted from it. In the past, Machine-Learning (ML) techniques have been widely used to capture hidden knowledge from stored historical data. In each case, the goal was to determine trends or behaviour patterns that would allow the improvement of KBS procedures. For instance, ML techniques have been used in hydroelectricity to produce rules from a power generation database [4] [5].

On the other hand, Object-Oriented (OO) approaches and languages have become quite popular, partially because of their potential benefits in terms of maintainability, reusability, separation of concerns, information hiding, etc. However, the vast majority of software available today is not OO. The effort to simply rewrite them from scratch using an OO approach would be prohibitive, and significant expertise recorded in the procedural software would be lost. The cost of manual conversion would also be prohibitive. Support coming from tools, documentation, and developers of the legacy software would ease the introduction of OO technology in many organizations. This kind of reengineering process could be especially helpful to integrate existing systems and new ones developed with OO approaches.

In the balance of this paper, we first present in section 2 the description of the context in which this work takes place, and the motivations for moving the Alcan Ltd legacy system towards an object oriented knowledge-based architecture. In section 3, we describe the adopted architecture and present its main features. Section 4 is devoted to present the machine-learning framework, which is part of the global system, and its performances. Finally in section 5, we conclude and present some of the lessons we learned.

II. THE DESCRIPTION OF THE CONTEXT

Alcan is one of the two world's biggest players in the aluminum industry. With a total surface area of 73 800 km^2, the Alcan hydropower network under study, constitutes a territory larger than the province of New Brunswick (Canada). The network has, on average, an annual energy capacity of approximately 2000 megawatts; it includes 6 hydroelectric power stations, 28 reserve installations, 43 turbine-alternators groups (TAG), 850 kilometers of energy transport lines, a network of about thirty hydro-meteorological stations, etc. Figure 1 illustrates the geographical localization where the work takes place.

---

*This work is part of an industrial project between the ALCAN Ltd group and CRIM, a research center. It was supported by a joint grant from ALCAN Ltd and NSERC, operation grant #CRDPJ 228746-99.

The objective of planning the operation of such a network can be summarized as the satisfaction of the following requirements:

- Effective use of water
- Account of future hydrological uncertainty
- Satisfaction of energy need
- Respect of safety constraints.

To reach these goals, a decision-making process of water stock management is used that consists of four steps:

1) Weather hydro measurements and gathering of the data;
2) Data analysis;
3) Weather and hydrological forecasting;
4) Planning.

In these planning tasks, information processing systems based on mathematical models tested for this kind of applications, are used for optimization and simulation purposes. These models are implemented in Fortran within more than 65 routines. A part of the application contains what we consider as expert knowledge within its source code. However, most of the knowledge is used implicitly and in a non-automated way by Alcan analysts at different steps of their decision-making process. Thus, the main disadvantage of the solution used since several years, is the absence of separation between the knowledge level and the inference or reasoning one, in a product used in a knowledge intensive process! An immediate consequence is the restriction in the possibilities of investigation and exploration wanted by the Alcan analysts. A consequence of that is the difficulty of maintaining and making evolve such a system.

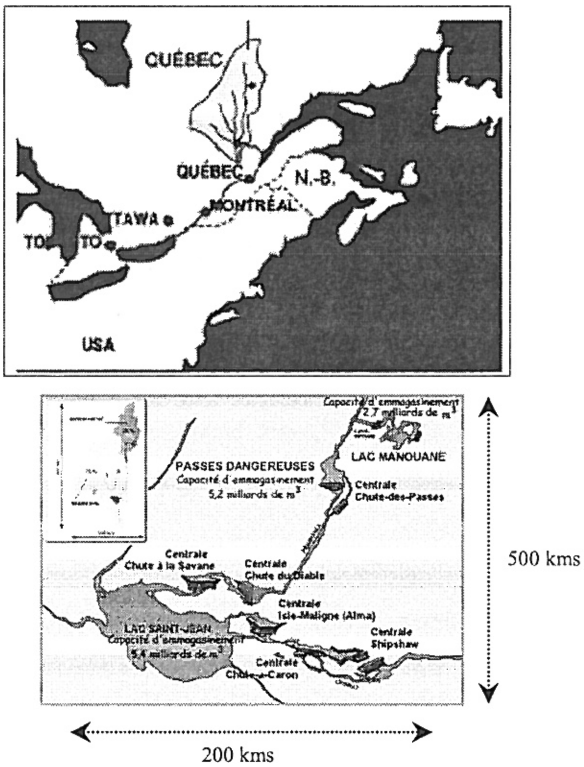

Fig. 1. Geographical localization of the study

The decision-making process is part of the knowledge management (KM) policy of Alcan Ltd. However, the knowledge used in the studied decision-making process, is many and varied. The problem is that this knowledge is often hidden in the code of the programs which use it, or consigned in internal documents, or even used implicitly by the experts, as in our case. This situation becomes more problematic, when the Alcan hydrological resources analysts wants to explore new scenarios, while modifying a little one of this knowledge. It has no other choice than to traverse the source code of the implemented programs, in order to make the discounted modifications there. It is, for example, the case of the operation rules of each power station and tank.

The available knowledge can be explicit or implicit. An explicit representation consists of a symbolic expression of human expert knowledge. Rules are an example of that; they allow you to separate the expertise from the application code. Since expert rules are externalized from the application code, they can be changed independently without recompiling the application. An implicit representation is knowledge that is usually hidden in data. It requires further processing of the data before useful information can be extracted from it. For that, Machine-Learning (ML) techniques have been widely used to capture hidden knowledge from stored historical data. In each case, the goal was to determine trends or behaviour patterns that would allow the improvement of KBS procedures. In our study, we want to improve the forecasts of natural contributions flow thanks to information contained in the historical database. Section 4 will introduce first the variables we want to predict and the predictive variables, and then, the machine-learning algorithms we have used to do so.

In the following section, we present the main features of the reengineered system.

III. THE OO KNOWLEDGE-BASED SOLUTION

As stated above, this work deals partly with the knowledge management of the Alcan experts, including data, thus allowing the hydropower resources planning or simulation. To help perform the planning tasks, we have developed a KBS called HYPERPIK (Hydro Power Resources Planning based on Inference and Knowledge). Figure 2 summarizes our system architecture.

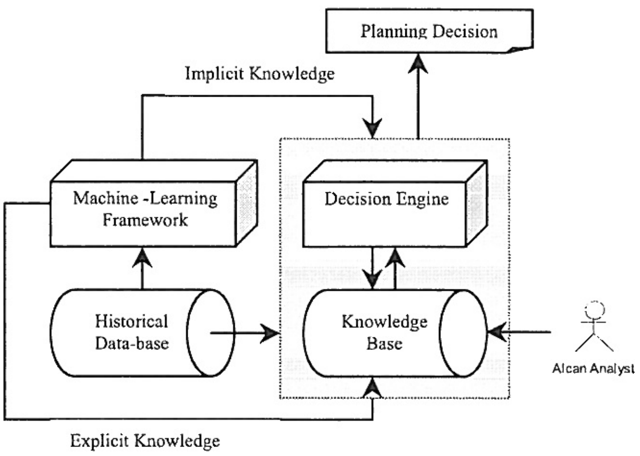

Fig. 2. The HYPERPIK architecture

It consists of an inference engine that is coupled with a knowledge base resulting from the problem modeling. The knowledge base contains an explicit knowledge that is the symbolic expression of Alcan experts' know-how. A machine-learning framework exploits a historical database and produces explicit or implicit knowledge, depending on the selected learning mechanism. The produced knowledge is then used in the decision process. In particular, it uses natural contributions flow values predicted from the historical database. These contributions flow values help evaluate the ability of the power system to face various contingencies and to propose appropriate remedial actions.

On the other hand, the planning step exploits explicit knowledge. It takes the form of rules. Rule technology is based on the philosophy of providing fast and flexible software components to empower computer applications with "business" or "expert" rules capabilities. The general idea of a rule is that actions on the right-hand side are carried out whenever all the patterns on the left-hand side are successfully matched. A *pattern* is an expression that is capable of designating one or more objects. The objects result from our modelization of the hydropower domain and figure 3 illustrates the resulted classes diagram.

The decision (inference engine) processes the rules using the objects in a working memory. It implements a RETE algorithm [6] (it is widely recognized as by far the most efficient algorithm for the implementation of production systems) where rules are compiled into a network. Input data to the network consists of changes to working memory. Objects are inserted, removed and modified. The network processes these changes and produces a new set of rules to be fired. This process continues cyclically until there are no further rules to be fired.

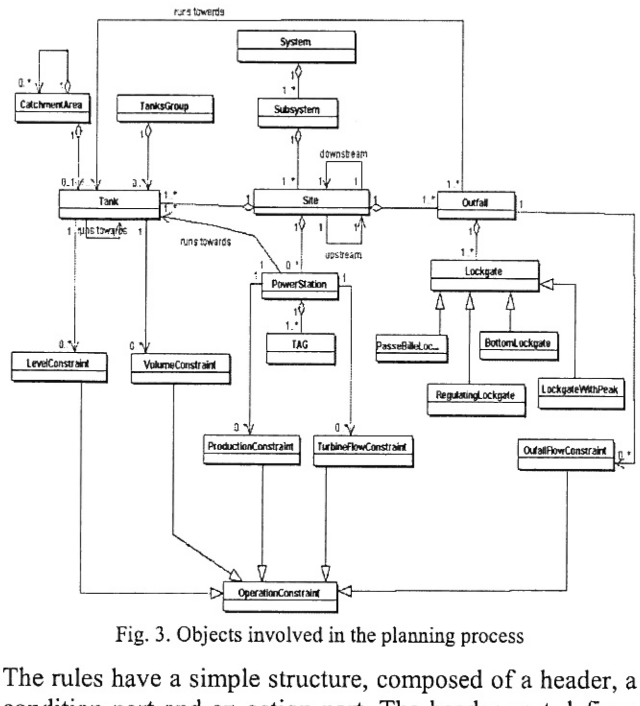

Fig. 3. Objects involved in the planning process

The rules have a simple structure, composed of a header, a condition part and an action part. The header part defines the name of the rule, the eventual packet to which the rule is attached, and its priority (if needed). The condition part utilizes the object-oriented structure of Java to carry out pattern matching on class instances, i.e. objects. This pattern matching binds (instantiates) variables to objects and field values. Rule conditions are also used to test field values. This provides a filtering mechanism for objects. When the condition part of a rule is verified, i.e. valid objects have been found, the action part of the rule may be executed. Actions may vary from simple to complex, e.g. printing a message to creating new objects or calling a pre-existing Fortran routine (through a Java method). The rules are written in the Ilog rule language[1] and the following one illustrate their structure:

rule
MaxEnergyProduction_StJeanLake_withoutDischarge{
packet = Management_StJeanLake;
priority = 10 ;
when { Simulation(hydriousContext(currentDate) == "Winter");
?lsj : Site (?lsj .name == "StJeanLake";
?lsj. shortTermRiskDischarge (currentDate) >= 0);}
then { //product Max energy at StJeanLake without discharging
modify ?lsj { ?lsj.turbineMaxWithoutDischarge (currentDate) ; } }}

Our work yields to a knowledge base of about 150 rules organized in 16 packets. A packet allows us to group rules with regard to their goal in the whole process. Examples of packets are: St-Jean Lake management (see the rule above), short-term risk at St-Jean Lake, overflow risk, Saguenay sub-system energy production, etc. Some of them exploit a priority to determine the order in which rules are executed. The larger the number is, the higher the firing priority of the rule.

A new interesting functionality of the new system concerns its evolvability. The Alcan analysts can edit the knowledge base and then modify one or several rules before running a new simulation and then exploring new scenarios. It is one of the main requirements that motivate this work, and figure 4 is an illustration of this new functionality.

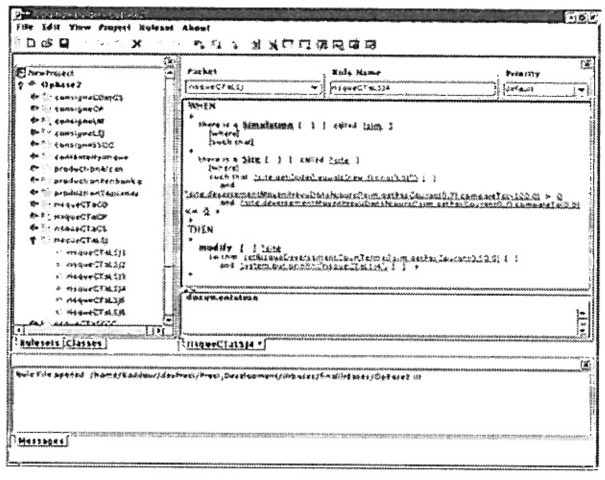

Fig. 4. The rule base editor

[1] Ilog Jrules is a general-purpose expert-system generator that combines rule-based techniques and object-oriented programming (www.ilog.com).

This reengineering work yields also to a new interaction model between Alcan analysts (final users of the system) and the system. The use-cases given in figure 5 illustrate these new functionalities and particularly the flexible way that the experts from now on have to configure their network or run a simulation session. Thus, an analyst can configure (define) the hydropower network; it means defining the whole system and all the elements it contains (e.g., power stations, tanks, etc.). Of course, he can run a simulation to obtain planning results; this simulation is done thanks to the rules we have implemented and also to some computations (forecasting) done by the Fortran routines. Figure 6 is an example of a simulation for a winter season.

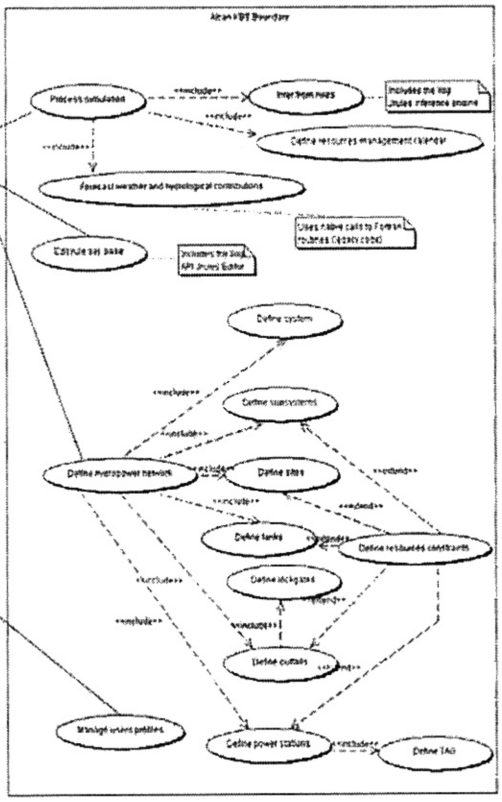

Fig. 5. The Reengineered system use-cases

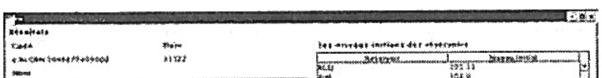

Fig. 6. Running a simulation

Next section will describe the machine-learning framework we are developing. It is an important piece of the whole architecture. It positively influences the evolvability of the system, by considering data in the planning process.

IV. THE MACHINE LEARNING FRAMEWORK

In our study, the machine-learning framework uses information contained in a historical database to improve the forecasts of natural contributions flow in the short run of Chute-du-Diable (CD – see the map in Figure 1) catchment area for the summer-fall period (June 15 at November 30). The variables we have worked on are summarized in tables 1 and 2, which list the variables to be predicted and the predictive variables.

TABLE 1. Variables to be predicted

Variable	Sample years	Abbreviation
Natural CD contributions flow for day 1 to 7 in a week	1992 – 99	AncCD_J1 to AncCD_J7
Volume of natural contributions flow of next 3 days	1992 – 99	Vol_3J
Volume of natural contributions flow of next 7 days	1992 – 99	Vol_7J

The predictive data includes the discharge flows of several rivers and serves to measure the natural contributions to a site, as these are the sum of the flows of the various rivers of the catchment area. For the CD site, the Manouane, Serpent and Petite Péribonka rivers are all parts of the CD catchment area. The data also includes precipitations, as the more it rains, the more the natural contributions are high. Due to problems encountered while using this first data set, we worked with Alcan's experts to produce a more efficient one, as described next.

First, we restricted our study to the years between 1992 and 1999, since data form the previous years were often fragmentary. The incomplete data were eliminated after consulting the experts. We also created new variables to summarize redundant information. Among these is a balanced precipitation that accounts for the precipitations of the 7 or 8 days previous to the day for which one wants to estimate the natural contributions. Next, we worked with Alcan's experts on trying to reduce the number of predictive variables by considering the characteristics of the Alcan network and by only keeping the relevant variables. For example, since the flow of the Manouane and Serpent rivers is measured upstream CD, one considers that the measured flow for the Manouane river at time T will arrive at CD at time T+24 hours, and that of the Serpent river at time T+18 hours. This shift thus enables us to have a more precise idea of the flow coming from these two rivers, which will contribute to the natural contributions of the following day. Since the two rivers do not cover the whole CD catchment area, the discharges of Petite Péribonka river are also needed to partially account for the flow of the southern part of the catchment area.

Thus and from the above considerations, we obtained the refined predictive variables given in table 2. These were the starting point for selecting the relevant variables to use in predicting the values of natural contributions from historical data. The selection process was done with the help of Alcan's experts, based on relevance and the experts' experience. It turned out that 14 to 16 of the 51 predictive variables were needed for each variable to predict.

TABLE 2. Refined predictive variables

Variables	Sample years	Number of data	Abbreviation
Natural contribution flow for previous 2 days at CD	1992 – 1999	2	AncCD_P1 AncCD_P2
Variation of natural contribution flow between Day –1 and Day –2 at CD	1992 – 1999	1	AncCD_T1j
Discharge flow of Manouane river at 4H	1992 – 1999	1	Qman_4H
Discharge flow of Manouane river previous day	1992 – 1999	1	Qman_P1
Variation of discharge flow of Manouane river for previous 4H, 12H, 24H (m3/s / hr)	1992 – 1999	3	Qman_T4h, Qman_T12h, Qman_T24h,
Variation trend of discharge flow of Manouane river between 4 and 12 hours	1992 – 1999	1	Qman_VarT
Discharge flow of Serpent river at 4H	1992 – 1999	1	Qserp_4h
Discharge flow of Serpent river previous day	1992 – 1999	1	Qserp_P1
Variation of discharge flow of Serpent river for previous 4H, 12H, 24H (m3/s / hr)	1992 – 1999	3	Qserp_T4h, Qserp_T12h, Qserp_T24h
Variation trend of discharge flow of Serpent river between 4 and 12 hours	1992 – 1999	1	Qserp_VarT
Discharge flow of Petite Péribonka river at 4h	1992 – 1999	1	Qpper_4h
Discharge flow of Petite Péribonka river previous day	1992 – 1999	1	Qpper_P1
Variation of discharge flow of Petite Péribonka river for previous 4H, 12H, 24H (m3/s / hr)	1992 – 1999	3	Qpper_T4h, Qpper_T12h, Qpper_T24h
Variation trend of discharge flow of Petite Péribonka river between 4 and 12 H	1992 – 1999	1	Qpper_VarT
Precipitation from day –4 to day –1 - catchment area CD	1992 – 1999	4	PbvCD_P1 to PbvCD_P3
Precipitation from day –3 to day –1 – station area CD	1992 – 1999	3	PstCD_P1 to PstCD_P3
Precipitation from day –3 to day –1 – station area CDP	1992 – 1999	3	PstCDP_P1 to PstCDP_P3
Precipitation from day –3 to day –1 – station area Mistassibi 2	1992 – 1999	3	PstMisbi2_P1 to PstMisbi2_P3
Precipitation from day –3 to day –1 – station area Manouane Est	1992 – 1999	3	PstManE_P1, to PstManE_P3
Precipitation forecasts from day+1 to day+7 – catchment area CD	1992 – 1999	7	Pprev_J1 to Pprev_J7
Balanced precipitation forecasts from day+1 to day+7 – catchment area CD	1992 – 1999	7	Ppond_J1 to Ppond_J7

Starting from the taxonomy given in [7] and illustrated by figure 7, we have explored four ML algorithms to predict the variables needed in the planning process: C4.5 [8] a top-down induction decision tree algorithm, CN2 [9] [10] a rule induction algorithm, a Case-Based Learning (CBL) algorithm, and a resilient back-propagation (RPROP) neural network [11] [12].

In this work, the set of refined historical data, about 1200 values for each predictive variable, was randomly divided in two subsets with two thirds of the data in the first one, and the remainder in the second. The first subset was used for training the network and the second for testing it.

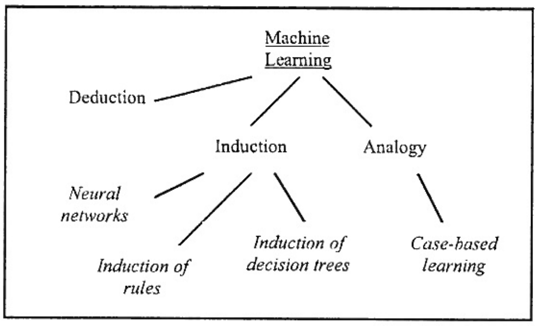

Fig. 7. Machine-Learning Taxonomy

We also used a cross-validation procedure where the available data is divided into N blocks so as to make each block's number of cases and class distribution as uniform as possible. N different classification models were then built, with one block omitted from the training data in each model. Each model was then tested on cases that belong to the omitted block. In this way, each case appears in exactly one test set. Provided that N is not too small, the average error rate over the N unseen test sets is a good predictor of the error rate of a model built from all the data.

Table 3 shows the best results that we have obtained by using the different ML methods.

TABLE 3. Computed Accuracies (%)

Algorithms / Variables	C4.5	CN2	CBL	RPROP NN
Vol_3J	91	91.5	99.3	99.1
Vol_7J	77.1	84	99.6	97.2
Anc_CD_1J	79.7	80.3	97.9	94.2
Anc_CD_2J	81.4	80.7	96.2	95.7
Anc_CD_3J	77.1	80.9	98.3	94.4
Anc_CD_4J	77.7	76.6	95.5	94.6
Anc_CD_5J	80.3	76.6	96.5	95.2
Anc_CD_6J	82.3	85.1	99.3	93.3
Anc_CD_7J	81.4	78.7	99.3	93.4

The computed accuracies are pretty high especially, for the Vol_3J variable. The predictive power of the different models is very satisfactory!

One distinction between C4.5 and CN2 from one part, and RPROP and CBL from the other part, is on the induced knowledge. The two first produce rules or decision trees that could be exploited by the knowledge-based system in its decision making process. RPROP and CBL do not produce any kind of explicit knowledge. However, they allow predicting variables like volumes or contribution flows, giving some predictive variables.

In the light of the obtained results, it seems that the most efficient way to build the ML framework is to associate the multilayer perceptron or the CBL algorithm with one of the inductive algorithms, C4.5 or CN2. The former would generate implicit knowledge (variable classification) that helps the decision process and the latter would generate explicit knowledge (rules) to be stored in the knowledge base.

V. CONCLUSIONS

The reengineered system is currently used within the hydropower resources management group at Alcan Ltd. It results from a long collaborative process between authors of this paper and Alcan analysts. The latter were active and decisive actors; they have to maintain their Fortran routines (e.g., short term evaluation functions, water rise rate calculation functions, volume calculation functions, etc.) and we have to build a bridge between these functions and the objects methods we have implemented. The exercise was not so easy; we have to keep a good separation between what we consider as an expert knowledge and the procedures that exploit this knowledge. This critical step of the project was iterative and it requires, even now, many adjustments.

The following points could summarize the strengths of this solution:

- a greater flexibility of the tool during its use within the decisional process, by facilitating the exploration of new power network management scenarios. It is the main need of Alcan analysts;
- better user interfaces allowing better usability during system configuration and simulations;
- a better evolvability of the system thanks to the OO rules-based architecture. It results from our objects/rules modelization and it allows updating easily expert knowledge to explore new planning schemes. The general understandability of the system is also much higher than it was in the previous system;
- a current and future better reusability of different components of the system.

We are working now on some extensions of the system. By tuning rule priority factors, mainly for those dealing with management instructions, we expect to improve the whole performance of the tool. In fact, the tuning of such a system is a long and meticulous work; it is actually one of the main tasks of Alcan analysts. On the other hand, we are still exploring new ML algorithms to incorporate within our framework. Finally, a next step will be to produce a rules verification module, coupled with the rule base editor, in order to maintain the knowledge base free of anomalies (redundancy, inconsistency, etc.).

VI. ACKNOWLEDGEMENT

The authors want to thank Ms. Louise Rémillard and Janine Dufour from Alcan Ltd for their precious collaboration. They also thank M. Boukadoum, V. Siveton and G. Maatoug for their collaboration in some aspects of the machine learning work.

VII. REFERENCES

[1] S. Samarasinghe, A. McKinnon, and J. Bright, "Expert System for Flood Management in Lake Manapouri ", IEEE Comput. Soc. Press, Los Alamitos, CA, USA. First New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems, 1993.

[2] J. Chang, Tiao and D. Moore, "Reservoir Operation by the Use of an Expert System", Proc. of the 1994 ASCE National Conference on Hydraulic Engineering. Buffalo, NY, USA.

[3] A.S. Leslie, A. Moyes, J.R. McDonald, G.M. Burt, J. McGowan, and W. Charlesworth, "CEPE, Intelligent System for the Management of a Hydro-electric Scheme", 31st Universities Power Engineering Conference, Iraklio, Greece, 1996.

[4] M. Mejia-Lavalla and G. Rodriguez-Ortiz, "Obtaining expert system rules using data mining tools from a power generation database", Expert systems with applications, 14, 1998, 37-42. 1517.

[5] H. Lounis, M. Boukadoum, and V. Siveton, "Assessing Hydro Power System Relevant Variables: a Comparison Between a Neural Network and Different Machine Learning approaches", Proc. International Conference on Neuro-Fuzzy Technology (Neuro-Fuzzy2002), Havana (Cuba), January 2002. 45-51.

[6] C.L. Forgy, "Rete: A Fast Algorithm for the Many Pattern/Many Object Pattern Match Problem", Artificial Intelligence, 19(1982) 17-37.

[7] Y. Kodratoff, "Apprentissage symbolique : une approche de l'IA", tome 1&2, Cépaduès-Editions, 1994.

[8] J.R. Quinlan, "C4.5: Programs for Machine Learning", Morgan Kaufmann Publishers, 1993.

[9] P. Clark and T. Niblett, "The CN2 induction algorithm", in ML Journal, 3(4):261-283, 1989.

[10] P. Clark and R. Boswell. "Rule induction with CN2: Some recent improvements", in Y. Kodratoff editor, Machine Learning - EWSL-91, pages 151-163, Berlin, 1991. Springer-Verlag.

[11] D. E. Rumelhart and J. L. McClelland, "Parallel Distributed Processing, Explorations in the Microstructure of Cognition", vol. 1: Foundations, MIT press, 1986.

[12] M. Riedmiller and H. Braun, "A Direct Adaptive Method for Faster Back-propagation Learning: The RPROP Algorithm", Proc. of the IEEE Intl. Conf. on Neural Networks, San Francisco, CA, 1993.