Epiphenomenal Intelligence from Partial Models in Safety Management

Marco Lazzari

Faculty of Arts and Philosophy
University of Bergamo, Italy

Abstract

This paper represents an ex post rethinking of the contribution of artificial intelligence techniques to safety management, based on a long work experience in applying artificial intelligence to several engineering fields, from dam safety to environmental protection, from seismic monitoring to the protection of cultural heritage. The main issue is that developing models for assessing safety is a hard task, but integrating partial models may provide good results. Even if each single model is poor, a sort of epiphenomenal intelligence emerges from the behaviour of a system made of small partial models and users perceive it as a reliable assistant.

Keywords: artificial intelligence, safety management, epiphenomenal intelligence, monitoring, ethnography, anthropology

1. Decision support systems and safety management

This paper represents an ex post rethinking of the contribution of artificial intelligence techniques to engineering and environmental safety management, based on a 13-years long work experience [1], set up when the author of this paper served as a software engineer at ISMES, a private research and development institute involved with safety management in several engineering fields, from dam safety to environmental protection, from seismic monitoring to the protection of cultural heritage.

In its 50 years long life ISMES developed outstanding research in safety management and developed techniques and systems to deal with different problems in the field of safety management, such as seismic vulnerability assessment of buildings, dam safety, monitoring of landslides and seismic and volcanic regions; these systems, comprised of monitoring instruments, hardware and software, have been installed in hundreds of sites in Italy and abroad [2, 3, 4].

The experience acquired by managing emergency situations, such as those of the ruinous Valtellina landslide or the collapse of the Civic Tower in Pavia, taught the lesson that the emergency must be tackled through the creation of an organisation that manages the knowledge of the actual conditions and their evolution as provided not only by basic studies, but also by instrumentation networks that monitor the most critical phenomena.

In this field artificial intelligence concepts and technologies can assist engineers and safety managers by providing additional software components to be integrated into existing information and monitoring systems, which may perform *intelligent* processing of safety related data.

When we started our project aimed at applying *artificial intelligence* to *safety problems* in *structural and environmental engineering*, we were in the late eighties and we firmly trusted in the possibility of developing a traditional expert system able to master the domain and to drive its users from *monitoring data*, *visual inspections* or *test data* to safety evaluations.

Few years later we had changed our viewpoint: we were dealing with ill-structured domains, where the objects to be checked - *dams, monuments, landslides* - are one-off objects (cannot be tested or statistically defined), behave as a continuum, interact with the social environment, are subject to uncontrolled / unpredictable input, cannot be *controlled*, are difficult and expensive to be measured. And, eventually, their knowledge is often based on partial models.

All these characteristics led us in the early nineties to restrict our target to a family of *decision support systems for the real-time evaluation of monitoring data* which were essentially model-based interpretation systems made of several *partial models* of the system to be checked.

According to this choice, the aim of those support systems was to help safety managers, engineers and authorities to deal with safety problems of structures in two ways:

- by filtering false alarms;

- by supporting the automatic detection and early warning of dangerous situations.

The former is mainly related to the reduction of costs required by human interpretation in case of false alarms, whilst the latter deals with the improvement of the safety level of the structure.

From the organizational point of view, two kinds of situation are interesting for the use of such systems. The first one is the interpretation of data coming from structures of particular relevance or criticality. The second is the case of a significant amount of structures (e.g. a set of dams of a region) operated by a central office that collects all the data and monitors the status.

The partial models composing our support systems, co-operating via a blackboard architecture, were able to provide *partial interpretations of the phenomena going on*; the co-operation among them, which is often achieved by simply highlighting the most conservative deduction about the data, not necessarily gives to the users *the answer*, but provides them with reasonable interpretations, accurate filtering of false alarms and, finally, *increases the dependability of the system under surveillance*.

As a result, dangerous situations are promptly spotted whilst expenditures for expert interpretations of false alarms are dramatically reduced.

The models that we implemented were of different nature, depending on the type of knowledge available and gathered from experts: mathematical models of well understood and documented relations among sets of different measurements, qualitative causal models, empirical formulas, associational rules

As a result of the different kinds of instruments taken into account and of the different kind of knowledge (and importance) of their relation with safety, it may happen that some reasoning paths *from data to interpretation* are made of single checks of measures against thresholds, or arise from a function that takes into account several dozens of instruments together with their coefficients of relevance and reliability, or derive from the processing of the whole set of hourly measures of an instrument recorded during one year.

Although rather satisfied by our achievements, we did explore other solutions, often pressed by *specific users' requirements* arisen from the everyday use of our support systems.

This led us to implement some *task specific agents* to solve specific problems [5], exploiting different techniques, from case-based reasoning to neural nets, from tailored explanations to on-line graphical assistants.

2. Structure and functions of the DSS

The final result of the efforts described in the previous paragraphs is a kind of decision support system (DSS) – already installed on six large dams [6], seven monumental structures [7], a dozen of landslides [8] - that embodies several different partial models of the knowledge about the object to be assessed, implemented as shortly described in the following:

- *numerical pre-processing*
  A numerical pre-processor checks the measurements and their rate of change against thresholds; moreover, it exploits theoretical models for computing the expected values for (some of) the instruments, then evaluates the displacements of the measured values from the expected ones and their rate of change and checks them against safety thresholds. The thresholds used within the DSS have a different meaning from that used by the acquisition system when the two systems are coupled: whilst the former are safety thresholds,

individuated by safety experts on the ground of their knowledge of the structure and of the past measurements, the latter are purely regarded as instrumental thresholds, used to verify the correct behaviour of the instruments.

- *empirical and model-based interpretation*
  The interpretation is a process of evidential reasoning, which maps data states into alarm states. The interpretation may be based on three types of knowledge:

  1. *empirical knowledge* - states of sensors, derived from comparisons between measurements and thresholds, may be combined through sets of rules. The rules, based on empirical past experiences or derived from statistical processing of past measurements, may be applied to groups of instruments of the same type, providing a synthetic index having the meaning of an anxiety status. Rules may be also applied to sensors of different type belonging to the same part of the physical system (e.g. a dam block). These rules may highlight the existence – and trace the evolution - of dangerous phenomena in that part of the physical system

  2. *qualitative models* - qualitative relationships between measured variables may be evaluated to provide evidences for physical processes (e.g. excessive rotation of a dam block, change in seepage around the dam, opening of a crack)

  3. *quantitative models* - quantitative models may be used, in co-operation with thresholds, to provide states of sensors; these models implement relations between cause and effect variables based either on statistical processing of past data or on deterministic modelling approaches

- *case-based interpretation*
  Given the qualitative description of the state of a structure, analogical reasoning is used to retrieve the closest-matching cases stored in a case base, which can help safety managers to interpret the current situation. Usually, situations stored in the reference database are those which were more deeply investigated by experts, because of some singularities, and are enriched with experts' comments; at run-time the analogical reasoning allows to understand whether the current situation is similar to a past one stored in the case base, and the experts' comments related to the latter can address the investigation of the current one. This enables to record and deal with special behaviours that cannot be adequately managed by explicit models.
  Two different approaches were adopted for developing these tools.
  The first approach is based on numerical/symbolic processing: several metrics were defined to compute the analogy of a couple of cases. These metrics span from simple norms within the hyperspace defined by the state indexes, to more complex functions also taking into account the gravity of the cases, in a fashion similar to the law of gravitation. In such way, instead of computing the distance between two situations, their reciprocal attraction is evaluated and checked against an adequate threshold system.

The second approach is based on the application of neural networks to implement a sort of associative memory that allows to compute the similarity of a couple of cases.

- *empirical interpretation based on neural networks*
  Within the DSS deep knowledge is usually codified by numerical algorithms, whilst qualitative reasoning is implemented by symbolic processing; when heuristic knowledge is involved, the shallow knowledge about the structure and the instrumentation is managed through empirical rules based on the alarm state of single instruments and on their reliability and significance. These rules are derived from the analysis of a set of exemplary cases, which allow to identify weights to be given to the parameters used by the rules.
  This process is time consuming and boring for the experts, but can be adequately dealt with neural nets. Therefore, neural nets can be developed to perform the empirical evaluation of data in order to achieve the same results as the symbolic processors, but with reduced development and tuning effort.

- *statistical processing*
  Statistical analysis of data may be performed both to provide evidences for the evaluation process, in order to spot trends in data (e.g., "the measure of a plumb line is constantly increasing"), and for periodical synthesis of past evaluations (automatic weekly reports or reports on demand).

3. Discussion

Each of the models introduced in the previous paragraphs is often poor and would not provide an answer good enough for safety management problems. Nevertheless, the appropriate behaviour is given by the simultaneous action of many different agents, their co-ordination via the connective structure of the system (a blackboard for exchanging data and partial results, a historical database for data and evaluations, the reference cases' database used to look for analogies, a GIS based interface for displaying data and alarm states), and the availability of appropriate links from on-the-fly generated explanation texts to the hypertextual documentation about the monitored objects.

Therefore, a form of epiphenomenal intelligence [9, 10] emerges from the behaviour of a system made of small partial models: *users perceive the system as a reliable assistant*, able to filter false alarms, catch significant events, explain what is going on and which are the causes and the possible effects.

Moreover, sometimes the system provides answers which are not immediately regarded as valid by the safety experts, who only after a whole process of deconstruction of the solution suggested by the system agrees with it. This means that these results take systems' designers by surprise, since they reveal some forms of complex "intelligent" behaviour that "innocently emerges" [11] and that they did not explicitly put into the software.

The validity of the evaluations is confirmed by all the analyses that we performed off line to check the behaviour, precision and robustness of our systems.

For instance, an analysis of the results of the interpretation system which deals with the measurements of the Cathedral of Pavia, performed on about 850.000 measurements gathered from January 1994 to September 1996, pointed out that wrong data were properly filtered by congruency checks, avoiding incorrect signalling and anomalies of the instrumentation or significant processes were timely detected [7].

An other deep analysis was conducted on a set of about 37,000 measurements managed by an interpretation system installed on an arch dam [1]. This instance of the DSS derives a judgement on the global state of the dam from row measures through empirical and model-based reasoning, taking into account numerical and statistical processing; case-based reasoning is not applied, since the customer did not require this functionality.

During the period taken into account the DSS highlighted a number of warning situations. But an off-line analysis of the data pointed out that the warnings were due to an extraordinary combination of events (higher level of the basin, abnormal air temperatures), which were outside the limits of the DSS's models. In fact, the result of the evaluations was, in those situations, a proper reminder for the dam managers of the unusual, delicate, but not severe operating conditions.

In the same period, a deep examination of the state indexes highlighted that the DSS classified as normal situations about 200 acquisitions that presented accidental over-threshold values. According to safety experts, such *filtering* was proper and correct; as a result, the dam managers were not requested to react to those situations, with a sensible reduction of management efforts.

On the other hand, when our systems highlighted some unusual situations, the off-line analysis proved that their judgement was correct, and those situations, although not dangerous, were suitable to be carefully taken into consideration.

The combined result of the behaviours is that the managers of the structures that we monitor are not distracted by false alarms, and can concentrate their attention and efforts on really significant situations. Moreover, the reduced number of stimuli avoids the *Oh, God, an other silly alarm* syndrome, which is really dangerous for safety managers, since it can induce them not to take into account warnings at all, even in severe conditions [12].

An additional support to the effectiveness of the DSS is given by its human-computer interface, specifically studied to take into account the different kinds of users of the system, according to principles of the user centered design.

A window-based interface draws on the screen graphical representations of the physical system and its composing objects as well as the sensors on the structure and

the interpretation objects. The interface displays the objects using a colour scale based on the objects' state; natural language explanations of the analysis, formatted through HTML, are shown on the screen too. Interactors are available to give users more refined information by focusing on interesting details, to activate subprograms for charting or reporting, or to access the internal data bases.

Moreover, when the monitoring instruments are spread over a large area, as for environmental monitoring, a window of the system may interface a GIS, to exploit its cartographic facilities and offer a powerful representation tool to *navigate* through data and exploit standard features of GISs, such as zooming functions to highlight sub-areas to be analysed.

An interesting side effect of our work is the collection and modelling of a corpus of knowledge, which was formerly belonging to a limited number of professionals. Even in the mind of those experts that knowledge was not organised in a clear conceptual structure suitable to be communicated, whilst now it is an explicit property of the company and may be reused, distributed and sold.

In a similar way, when we had to face the problem of supporting early age concrete crack prediction [13], we developed a knowledge based system simulating basic early age transient concrete properties, that embodies knowledge gathered from experts both in procedural form (programs and spreadsheets) and declarative form (graphical assistants, users' guides, documentation, database of concrete mixes), and where some processing tasks are performed by neural networks trained on data gathered from experimental and virtual models.

Even in this case the coexistence and integration of several different sources of knowledge provides users of the support they need to solve their problems and the coordinated behaviour of the different models which compose the system gives users the feeling of interacting with a rather clever device.

4. References

[1] Salvaneschi, P., Lazzari, M., "Integrating databases, data communication and artificial intelligence for applications in systems monitoring and safety problems", in "Database and data communication network systems", Leondes, C. T., (Editor), Academic Press, San Diego, CA, USA, 2002.

[2] Bonaldi, P., Carradori, G., Fanelli, M., Giuseppetti, G., Ruggeri, G., "Modern Techniques for Dam Safety Surveillance and Evaluation", in "Proceedings of the 16th ICOLD Congress", San Francisco, CA, USA, 1988.

[3] Anesa, F., Bonzi, A., Laquintana, D., Pilenga, A., Vavassori, M., "Valtellina alert system: towards an environmental risk diagnosis expert system", in "Proceedings of the IABSE Colloquium on Expert Systems in Civil Engineering", Bergamo, Italy, 347-354, 1989.

[4] Azzoni, A., Chiesa, S., Frassoni, A., Govi, M., "The Valpola landslide", "Engineering Geology", 33, 59-70, 1992.

[5]   Wielinga, B. J., Breuker, J. A., "Models of Expertise", in "Advances in Artificial Intelligence II", du Boulay, B., Hogg, D., Steels, L., (Editors), Elsevier Science, Amsterdam, The Netherlands, 1987.

[6]   Salvaneschi, P., Lazzari, "Applying AI to structural safety monitoring and evaluation", IEEE Expert, 11(4), 24-34, 1996.

[7]   Lancini, S., Lazzari, M., Masera, A., Salvaneschi, P., "Diagnosing of ancient monuments with expert software", Structural Engineering International, 7(4), 288-291, 1997.

[8]   Lazzari, M., Salvaneschi, P., "Embedding a geographic information system in a support decision system for landslide hazard monitoring", International Journal of Natural Hazards, 20(1), 185-195, 1999.

[9]   Hofstadter, D., "Gödel, Escher, Bach: an eternal golden braid", Basic Books, New York, NY, USA, 1979.

[10] Hofstadter, D., "Fluid concepts and creative analogies", Basic Books, New York, NY, USA, 1995.

[11] Dennett, D. C., "Brainstorms: philosophical essays on mind and psychology", Bradford Books, Montgomery, VT, USA, 1978.

[12] Mazzolini, L. A. R., "Dealing with uncertainty in geoepigraphic reasoning", Zenputer International, Menlo Park, CA, USA, 1997.

[13] Lazzari, M., Pellegrini, R., Dalmagioni, P., Emborg, M., "Integrating neural networks, databases and numerical software for managing early age concrete crack prediction", in "System based vision for strategic and creative design", Bontempi, F., (Editor) , Balkema Publishers, Lisse, The Netherlands, 2003