Environmental Risk Assessment as a Case-Based Preference Elicitation Process

Paolo Avesani and **Sara Ferrari** and **Angelo Susi**
ITC-IRST,
Via Sommarive 18 - Loc. Pantè, I-38050 Povo, Trento, Italy
{*avesani,sferrari,susi*}*@itc.it*

Abstract

The traditional approaches to environmental risk analysis based on first principle methods don't cover the assessment of systemic vulnerability, a meaningful component of the natural hazard. We argue that this complex task can be accomplished with a case-based approach through a process of pairwise preference elicitation. Up to now the main restriction of the case-base risk assessment was related to the scalability issue and the cognitive overload of the experts. In this paper we propose a methodology based on a mixed initiative strategy that combines the user preference elicitation and the machine rank approximation. The work includes both the case-based model and the related computational tools. We illustrate how a boosting algorithm can effectively estimate pairwise preferences and reduce the effort of the elicitation process. Experimental results, both on artificial data and a real world problem in the domain of civil defence, showed that a good trade-off can be achieved between the accuracy of the estimated preferences, and the elicitation effort of the end user.

1 Introduction

Environmental planning and more specifically the emergency planning for civil defence are configured as decision processes based on risk assessment. An environmental decision support system has very often to deal with the issues related to the natural hazards. The problem of risk assessment is usually not accomplished as a boolean decision procedure: yes or no? More often the risk analysis has to be designed as a ranking problem, i.e. finding a order relation over a set of alternatives [Roberts, 1979]. To elicitate a order relation is a knowledge intensive process because it requires to combine multiple points of view [Beinat and Nijkamp, 1998]. The challenge of the risk assessment is to enable the emergence of the critical factors through a context sensitive analysis of the real situations.

¿From the methodological point of view the complexity of this task is simplified to allow the exploitation of decision making techniques: definition of an utility function, collection of the preferences, synthesis of the rank over the set of alternatives. But the environmental risk assessment can not be easily encoded through an utility function. The complexity is most of the time underestimated because it is neglected one of the two components associated to the natural hazards: the systemic vulnerability. The risk analysis is traditionally concerned with the environmental vulnerability that refers to the magnitude of the disaster from the point of view of the physical phenomenons [E. J. Henley, 1992]. The systemic vulnerability refers to the component of natural hazard related to the human response to the disasters. The emergency organizations play a key role in determining the amount of damages: a quick intervention can reduce drammatically the number of victims and the impact on the commercial activities.

Although the physical phenomenons, like landslides or floods, are difficult to predict, there are powerful mathematical tools to describe these events. Thanks to simulations is possible for example to estimate in advance a quite accurate amount of the terrain volume moved by a landslide, or in the case of floods it is possible to know the territory that will be covered by the water. Although it seems counter intuitive it is very hard, even impossible, to estimate the delay of the rescue operations or the impact on the phenomenon due to the human intervention. Moreover it is very complex to elicitate in advance all the parameters that affect the role of the organization in reducing the natural hazards. For this reason the systemic component of the environmental risk assessment is usually neglected.

The goal of this work is the development of a methodology to allow a global risk assessment that will take into account both the environmental and the systemic vulnerability. Taking advantage of a case-based approach we will propose a method to support the elicitation of the relevant parameters and the related values for a specific situation. A further goal is the design and the development of the computational tools that will get effective the proposed methodology.

The objective of the methodology is to enable a complex analysis through a mixed initiative strategy that combines the knowledge of the expert with the scalable capabilities of the computer. Both of them should allow to achieve a context sensitive risk assessment. Furthermore we had the ambition to pursue two conflicting objectives: dealing with scalability issues and reducing the end user effort. It means to increase the number of risky areas that has to be studied without asking to the user a proportional overload. Of course the quality of the analysis will have to be not affected by the lower effort of the end user. These goals have been achieved exploiting the machine learning techniques that have taken advantages from the iterative mixed initiative strategy of interaction between the computer and the end user [Linden *et al.*, 1994]. An extended off-line experimentation has been performed to assess in advance the properties of the proposed model and to tune the computational tools. Finally a test on the field has been organized with real users and with a real world problem deploying a working system.

2 Motivations

The approaches to the environmental risk assessment can be divided into two main categories: *ex-ante* and *ex-post*. The former category refers to methods where the evaluation criterias are defined in advance, before knowing the specific instance of the problem; the latter category refers to methods where the evaluation criteria are defined after considering the specific instance of the problem.

The *ex-ante* approaches are based on first principle methods. The rationality of the assessment is encoded in a formal way, usually in terms of an utility function, without taking care of the data. For example in the domain of hydrogeological risk the criteria is defined by an equation that combines three concepts: danger, vulnerability and value ($R = H \cdot V \cdot v$). The problem is then decomposed into three independent subproblems: to estimate the probability of the events (H), to estimate the amount of the threatened goods (V), to estimate the values of the damaged goods (v). The recognition of these three kinds of information is collected from the analysis of the territory generating for each of them a layer. The overlapping of these three layers is obtained automatically with the support of the computer: the output is a new layer, the risk map like in Figure 1. This method is called *overlay model* and in the last years has been widely used thanks to the advent of the geographical information systems (GIS) that allow a large scale assessment. In spite of the fact that this tools overcome the scalability issue, a set of drawbacks prevents to perform an extended assessment of the risk including the systemic vulnerability [Carrara, 1991]. The main drawback is related to the collection of data, i.e. the acquisition of the single overlays. In this case there is the tacit assumption that, for example the *value* overlay, can be easily acquired without taking into account the context sensitive dependency of values.

Although the overlay model is successful in terms of scalability this method not necessary produces a total order over a set of risky areas. A partial order is not satisfactory when the risk assessment aims to support a decision making task. For example a civile defence plan includes a restricted budget to build artifacts that will reduce the risk; in such a case a generic categorization of the areas as highly risky is not effective to support a policy of intervention. The encoding of a first principle in advance prevents to know what will be the response on a given set of risky areas and consequently to tune a fine grained equation for risk analysis. But even with a fine grained model there is the strong preconditions that all the areas have to be described with respect to a common homogeneous schema. This premise is very restrictive and prevents to develop a context sensitive characterization of the single area: it is easy to understand that a flood and a landslide will be described using different parameters although both of them are concerned with the hydrogeological risk. Last but not least the first principle methods assume that it is possible to design in advance the effective schema to characterize a risky area. This assumption is usually optimistic and not sustainable when it is required to deal with the systemic vulnerability. Systemic vulnerability includes so many factors that a comprehensive schema would overwhelm any expert.

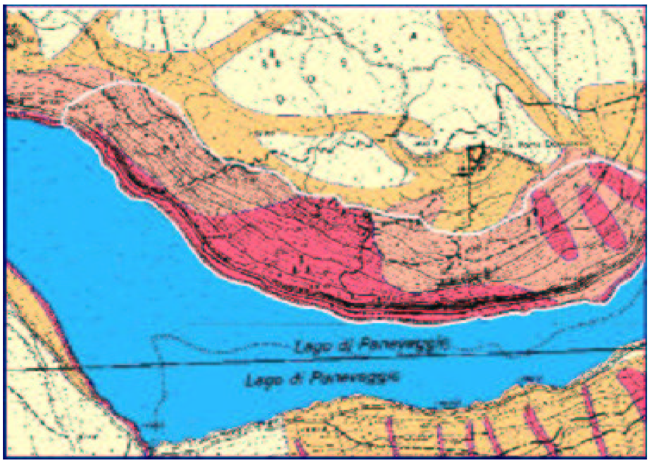

Figure 1: **The overlay model.** The visual output of a first principle approach based on Geographical Information Systems (GIS).

Overlay models and more in general first principle methods have a counterpart in *ex-post* approaches. In this case there is not an explicit encoding of the rationality underlying the risk analysis rather than an extensional elicitation of the risk values over a finite set of candidate areas. The elicitation of the risk value is encoded by a preference structure obtained through a contrast analysis. The risk value is not formulated with respect to an absolute range of values but as relative preference between two alternatives. Given two risky areas A and B the output will not be a categorization like A is low risk and B is medium risk but for example that A is more risky than B. Of course this way of proceed requires an exhaustive analysis of all the pairs of candidate areas. Given the complete preference structure over the set of all possible pairs a total rank of the areas is trivially computed.

This method is time consuming and labour intensive for the experts that are involved in an iterative process with a meaningful cognitive effort. The main drawback is that the amount of pairwise comparisons grows faster than the number of areas: the cardinality of the pairs set is quadratic with respect to the number of areas although we assume a symmetrical preference structure that reduces to an half the number of pairs the need to be taken into account. A symmetrical preference structure means that if we have two areas A and

B and the related preference relation $A \prec B$ the preference relation $B \succ A$ trivially follows.

The objective of this work is to enhance this method developing computational techniques that enable a reduction of the elicitation effort. The strategy that we will adopt is to decompose the problem into two parts: explicit versus implicit preference elicitation. The explicit preference elicitation involves directly the end user while the implicit preference elicitation is in charge of the computer. The implicit preference elicitation is accomplished through an approximation process that aims to learn the preference structure of the end user. Explicit and implicit preferences are combined together to obtain a global rank of the risky areas. In the following we will show how machine learning techniques can be effective in supporting the approximation step. Of course such a process is proned to errors. The challenge is to achieve an approximation error less than the human error latent in the huge elicitation process.

There is a further coordinated objective: supporting the partition between the explicit and implicit preference elicitation processes. It is to be noticed that the approximation step doesn't say anything on how to divide into two sets the set of pairwise preferences. For this purpose it is required a sampling policy to decide which pair preferences have to be acquired from the user and which ones have to be approximated from the machine. The two objectives, a preference learner and a sampling policy, can be summarized in a high level goal that aims to achieve a good trade-off between the elicitation effort and the rank accuracy.

3 Related Work

The case-based approach to risk assessment is well known in the scientific community overall because it has been formalized in a methodology called Analitic Hierarchy Process (AHP) [Saaty and Vargas, 2000; Saaty, 1999]. AHP is often used to perform environmental impact analysis where the main issue is to combine multiple viewpoints [Saaty and Vargas, 1994]. The strenghtness of AHP is that it doesn't require an explicit encoding of the preference rationality, it doesn't require an homogeneous representation of the cases, it support a case-based oriented assessment and consequently a context aware risk analysis. Let us briefly resume how AHP works.

The first step of AHP method is the selection of a given set of cases, called also *alternatives*. The second step is the enumeration of the criterias that have to be taken into account in the process of risk assessment. For each of them it is built a matrix where both rows and columns represent the set of areas under investigation. The objective is to fill half of the entries of the matrix (see the hypothesis of simmetrical preference relation mentioned above). Each cell is assigned with an integer number taken from the interval $[1, 9]$. The meaning of the number refers to a qualitative degree of the preference relation, for example that the area A is "moderately more risky" than area B. After the completion of the matrix a total order is synthesized through the computation of a vector of weights that specifies the rank of any given area. The elicitation process is replicated for each criteria that has to be taken into account. The vectors of weights and the related ranks represent the different points of view according to the predefined criterias. The ultimate step is the synthesis of a global rank that summarize the different views. An analogous preference elicitation process is performed on a matrix where rows and columns represent the different criterias taken into account. In such a way a rank on the criteria is obtained to combine a weighted composition of the different order relations defined over the set of alternatives.

It is quickly to assess that even with a small set of areas, say 10, it is required to elicitate 45 pairwise preferences for each criteria. The idea to reduce the elicitation effort it is not new and it has been already developed techniques for this purpose. But all these initiatives share the common restriction on the maximum number of alternative cases that AHP identifies around ten. Our strategy based on approximation aims to relax this constraint although at the same time relaxing the requirement on accuracy. To underline how much is restrictive the threshold on the number of cases it worthwhile to remark that in the case study we will show in the following the number of areas is 50. Moreover the AHP method has an additional constraint: the rated preference elicitation. Two are the related drawbacks: the former is the cognitive overload to specify one value preference out of nine rates, the latter is the semantic ambiguity of the rating values. Values like "moderately", "high" or "very high" usually adopted to rate the pairwise preferences are context sensitive both because are interpreted differently from different users and because their use may be sensitive to the order of case presentation.

More recently other initiatives have been developed to support the implicit preference elicitation to reduce the user effort. Although it has been conceived in an other domain it is worthwhile to mention a work where implicit preferences are acquired by the detection of user behaviour [Blythe, 2002]. All the cases are presented to the user to allow a global view over the set of alternatives. In this case the pairwise approach is surpassed and the issue of a sampling policy doesn't apply but an additional requirement arises: the visualization of the cases. To get effective the browsing of the space of alternatives it is needed to highlight the feature values that enable an assessment of the diversity. To project in the visual space many dimensions is still an open issue and for this reason the approach above has a strong limitation on the number of information that can be concurrently displayed for all the cases. This restriction is not sustainable in the domain of environmental risk assessment where a case is often described by many tens of features. We will see in the following how in our case study the pairwise comparison involves the display of two maps, a visualization requirement that doesn't allow to scale up with more cases at the same time.

Another work investigates some techniques to support the preference elicitation by approximation [Ha and Haddawy, 1998]. The basic intuition relies on the paradigm of case-based reasoning [Aamodt and Plaza, 1994]: to retrieve a past similar order relation to generate an approximation for the current ranking problem. A simple learner is proposed based on a nearest neighbour rule and the related similarity metric. The similarity metrics in this case are concerned with preference structures that can be only partially defined. This approach can be effective at two levels: the criterias or the cases. Both alternatives don't apply to the scenario of case-based environmental risk assessment: in the first case because the criteria are not explicitly encoded, in the second case because it is not available a preference structure already defined over the same current set of areas. The notion of previous experience usually applies to other set of risky areas and the similarity refers to the preference structure and not to the areas.

A similar approach to the learning of order relations is described in [Freund *et al.*, 1998] where the emphasis is not on the preference elicitation rather than on the approximation issue. RankBoost, a classifier based on the boosting of ranking learners is proposed. In particular, the goal of RankBoost is discovering a total order relation for a given set of instances described with some features, given a set of ordering functions for these features and a target user ordering function represented as a binary preference relation between couples of instances. Although these techniques have not been developed to support the approximation of preference structures, we will show in the section 5 how we have applied them to support the problem of preference elicitation.

To summarize, we are interested in exploiting the machine learning techniques mentioned above to combine a solution for the twofold problem: supporting the pairwise elicitation selection and the pairwise preferences approximation. In the next section, we will explain the process that integrates the two steps in such a way that each one can benefit from the other.

4 A Case-Based Elicitation Model

The objectives of this work can be summarized in the following points:

- **Elicitation advisory process.** Designing a case-based decision support system to support the interactive elicitation of a preference structure over a finite set of cases.
- **Pairwise preference elicitation.** Supporting the elicitation of a boolean preference structure through a pairwise comparison of cases.
- **Preference acquisition policy.** Developing an acquisition policy to support the expert in the process of preference elicitation: i.e. recommendation of the pairs of cases tha have to be analyzed first.
- **Preference learning step.** Exploiting learning techniques to approximate the unknown pair preferences taking advantage of the previously elicitated preferences.
- **Error and effort minimization.** Achieving a good trade-off between the accuracy of approximated rank and the effort required to collect the explicit definition of the preference structure.
- **Mutual recursive benefit.** Improving the accuracy of rank approximation taking advantage of an effective acquisition policy, and at the same time improving the acquisition policy taking advantage of the by-product result of the learning step (i.e. the pairs where the learner fails).

Before introducing the interactive case-based preference elicitation model we briefly sketch the general architecture that underlies the generic process.

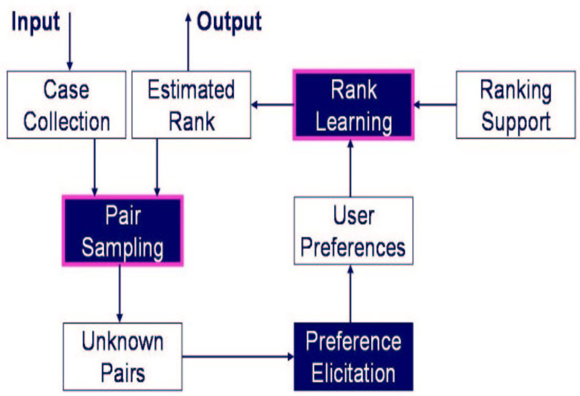

Figure 2: **The case-based model.** The white boxes represent the data while the black boxes represent the operations. Only the preference elicitation is in charge to end user.

The types of data involved in the elicitation process can be detailed as follows:

- data - The input data represent the finite collection of cases that has to be ranked. The data are the initial input of a case-based preference elicitation process.
- pair - A pair is a couple of alternative cases where the end user preference is unknown. Given a pair the end user is interactively asked for preference elicitation. The pairs are generated run time by an automated process.
- pref - The pref is the order relation between two alternative cases elicitated by the end user. The preference is formulated as a boolean choice from a pair, i.e. two alternative cases.
- ranks - The ranks are a collection of order relations defined on the finite collection of cases. When the ranks are derived by the feature-based description of the cases they are referred as *r*anking features.
- rank - The rank represents the target preference structure and it is the main output of the whole process. The rank produced in output could be an approximation of the exact rank.

Starting from this generic architecture, we can design at least four different models of the iterative process that enable the ranking of a set of cases through the interactive acquisition of pairwise preference: the basic single-user loop model, the iterated single-user loop model, the basic multi-user loop model and the self multi-user loop model.

The basic single-user loop model is depicted in Figure 2 and it is based on an iteration of three steps in the following sequence:

1. Pair sampling
   An automated procedure selects from the case base a

pair of cases whose relative preference is unknown according to a predefined acquisition policy.

2. Preference elicitation
   Given a pair of alternative cases the end user chooses which one is to be preferred with respect to the ultimate goal.
3. Rank learning
   Given a partial elicitation of the user preferences a learning algorithm produces an approximation of the unknown preferences and then a correspondent ranking of the whole set of cases is derived.

If the result of the learning step is considered enough accurate or the end user has been overloaded the iteration halts and the latest approximated rank is given as output; otherwise an other cycle of the loop is carried on. It is to be noticed that the first and the third steps are automated while the second step is in charge to the end user.

The basic single-user loop model is characterized by two properties: first only one user takes part to the elicitation process, second it is assumed that the preference elicitation is monotonic (i.e. the user doesn't see a pair twice).

In the iterated single-user loop model there is the assumption that the preference acquisition is context sensitive or better is history dependent. The choice between two alternative cases depends on what has been previously elicitated. Because the opinion of the users may change and they should have the opportunity to revise their preferences presenting to them the same pair of alternative cases.

The basic multi-user loop and the self multi-user loop models refer the scenario where many users take part to the elicitation process in order to take into account multiple points of view. More specifically the fourth model introduces a further restrictive condition: the learning step can't take advantage of the predefined order relations over the set of cases.

In the following we will restrict our attention to the basic single-user loop model.

5 Learning Pairwise Preferences

In this section we will focus our attention on the approximation of a preference structure. To explain how a boosting approach [Freund and Schapire, 1999] can be exploited for this purpose we arrange an abstraction of the problem definition.

The basic elements we have seen until now are the case base, the ranking features, the target rank and the elicitated pairwise preferences. The case base is defined as a finite set C of cases. The ranking features $F = (f_1, \ldots, f_N)$ are defined as a finite set of N features that describe the case, where $f_i : C \to \overline{\mathbb{R}}$ ($\overline{\mathbb{R}} = \mathbb{R} \cup \{\perp\}$) and the interpetation of the inequality $f(c_0) > f(c_1)$ means that c_0 is ranked above c_1 by f_i and $f_i(c) = \perp$ if c is unranked by f_i. The target rank is defined as a feedback function $\Phi : C \times C \to \mathbb{R}$ whose interpretation is that $\Phi(c_0, c_1)$ represents how important it is that c_1 be ranked above c_0; positive values means that c_1 should be ranked above c_0; $\Phi(c_0, c_1) = 0$ indicates that there is no preference between c_0 and c_1 (we assume $\Phi(c, c) = 0$ and $\Phi(c_0, c_1) = -\Phi(c_1, c_0)$ for all $c_0, c_1 \in C$). Finally the set of pairwise elicitated preferences is defined as $C_\Phi = \{c \in C | \exists c' \in C : \Phi(c, c') \neq 0\}$ that represents the feedback of the user or the finite support of Φ.

The goal of the learning step is to produce a ranking of all cases in C, including those not ranked by the f_i, represented in the form of a function $H : C \to \mathbb{R}$ with a similar interpretation to that of the ranking features (c_1 is ranked higher than c_0 by H if $H(c_1) > H(c_0)$).

The methodology we have choosen is boosting, a method to produce highly accurate prediction rules by combining many weak rules which may be moderately accurate; here we refer to the boosting algorithm introduced in [Freund *et al.*, 1998]. In the current setting the objective is a learning algorithm that will produce a function $H : C \to \mathbb{R}$ whose induced ordering of C will approximate the relative orderings encoded by the feedback function Φ using the information from the set of features F.

Before introducing the boosting algorithm we require to define an additional notion of *crucial pair*, a density function $D : C \times C \to \mathbb{R}$ such that

$$D(c_0, c_1) = \gamma \cdot max(\{0, \Phi(c_0, c_1)\})$$

setting to 0 all negative entries of Φ; γ is a positive constant chosen so that

$$\sum_{c_0, c_1} D(c_0, c_1) = 1$$

A pair c_0, c_1 is said to be *crucial* if $\Phi(c_0, c_1) > 0$, so that the pair receives non-zero weight under D. The algorithm we used is designed to find an order function H with a small weighted number of crucial-pair misorderings, in other words with a small ranking loss $rloss_D(H)$ defined as:

$$\begin{aligned} rloss_D(H) &= \sum_{c_0, c_1} D(c_0, c_1)[\![H(c_1) \leq H(c_0)]\!] \\ &= Pr_{(c_0, c_1) \sim D}[H(c_1) \leq H(c_0)] \end{aligned}$$

where $[\![H(c_1) \leq H(c_0)]\!] = 1$ if $H(c_1) \leq H(c_0)$ is true, 0 otherwise.

The algorithm *RankBoost* in Figure 3 operates in rounds. The basic round is based on two procedures called *WeakLearner* and *ChooseAlpha* that are invoked to produce respectively a *weak hypothesis* h and the value for the parameter α. *RankBoost* maintains a distribution D_t over $C \times C$ that is passed on round t to the weak learner. The algorithm chooses D_t to emphasize different parts of the training data. A high weight assigned to a pair of cases indicates a great importance and the weak learner has to order that pair correctly.

Weak hypotheses have the form $h_t : C \to \mathbb{R}$. The boosting algorithm uses the weak hypothesis h to update the distribution as shown in the algorithm. Suppose that c_0, c_1 is a crucial pair so that we want c_1 to be ranked higher than c_0 (in all other cases D_t is 0). Assuming that the parameter $\alpha_t > 0$ this rule decreases the weight $D_t(c_0, c_1)$ if h_t gives a correct ranking ($h_t(c_1) > h_t(c_0)$) an increases the weight otherwise. Thus, D_t tends to concentrate on the pairs whose relative ranking is hardest to determine.
The final hypothesis H is the α_t weighted sum of the weak hypothesis h_t.

Algorithm **RankBoost**
Input: initial distribution D over $C \times C$
Output: the final Hypothesis $H(c)$

begin
  Initialize $D_1 = D$;
  For $t = 1, \ldots, T$:
    $h_t = WeakLearner(.)$, where $h_t : C \to \mathbb{R}$;
    $\alpha_t = ChooseAlpha(.)$, where $\alpha_t \in \mathbb{R}$;
    $D_{t+1}(c_0, c_1) = \frac{D_t(c_0,c_1)}{Z_t} e^{\alpha_t(h_t(c_0)-h_t(c_1))}$;
  return $H(c) = \sum_{t=1}^{T} \alpha_t h_t(c)$;
end.

Figure 3: RankBoost algorithm.

During the development of the system we have implemented the weak learner proposed in [Freund *et al.*, 1998] that takes as input the ranking features $f_1, \ldots, f_N$ and the distribution D, and as output the learner h. Other possible weak learner schemas can be designed, for example the nearest neighbour learners based on the similarity metrics proposed in [Ha and Haddawy, 1998]; nevertheless combining nearest neighbour learners with a boosting schema produces classifiers more sensitive to the variance of the data.

6 Pairwise Sampling Policy

If the learning step illustrated in the previous section plays a key role in producing an accurate approximation of the target rank, an other task remains to be tackled: to reduce the elicitation effort of the end users. The purpose of the acquisition policy is to sample the pairs whose relative preferences are unknown. Of course if the number of elicitated preferences increases, the approximation error tends to decrease. When the acquisition process is exhaustive the approximation error is zero because all the information is elicitated. The challenge is to find a kind of minimum set that maximize the accuracy of the estimated rank. Although the model enables a one-shot pairs sample, the basic intuition is that an incremental strategy could be much more effective because after each learning step a new awareness on what is difficult to approximate is acquired. Therefore an acquisition policy can take advantage focusing the sampling on the pairs where the estimated preferences are not enough accurate [Caprile *et al.*, 2002].

As we have seen in the previous section the boosting technique produces, as a by-product of the learning algorithm, a distribution of weights that emphasizes the pair of cases for which it was harder to approximate the right preference. The rationality of acquisition policy combines this information with an other source of knowledge: the notion of problem complexity [Devetag and Warglien, 2002]. The complexity of the approximation problem can be characterized by the relationship between the target rank and the order relation used as the base of the inductive process. Four types of categories can be defined with increasing degree of complexity: *isotonic*, *antitonic*, *non-monotonic* and *non-projective*.

Although many acquisition policies can be designed all of them share the same structure based on two components: a boostrap component and a feedback component. The former samples a set of pairs without any hints because at the beginning no information are avaliable (remind that the sampling step is the first of the process), the latter implements the rationality that we sketched above.

In the experiments illustrated in the following section we have adopted a “bipartite” strategy for boostrapping the pair sampling. The hypothesis is to maximize the minimum coverage of the all cases: the pairs are generated without selecting the same case twice.

The next two sections are devoted to the experimentations of the model looking first at the evaluation setup and then at the empirical results.

7 Off-line Evaluation

The evaluation of the case-based preference elicitation model presented until now has been organized distinguishing between on-line and off-line experiments. The off-line experiments have been designed to assess the properties of the process and of the proposed solutions. The on-line experiments have been performed as a case study on a real world problem and with real users in the domain of the civil defence (see Section 8 for the details). As depicted in the Figure 2 all the experiments have tested the same three steps loop using the same acquisition policy and the same learning algorithm. Only the preference acquisition was different: in the on-line experiments the real users have been directly involved while in the off-line experiments they have been simulated.

For the off-line experiments many artificial datasets have been synthesized producing case bases with a number of instances ranging from ten to hundred. For each case base an arbitrary preference function Φ has been associated as the target rank and in a similar way the ranking features f_i have been derived from the case description. The Φ function has been used both to simulate the real user in the preference acquisition step and to assess the accuracy of the approximated rank function H.

The assessment of the accuracy has been performed using the notion of disagreement [Cohen *et al.*, 1998] between the two preference functions Φ and H: i.e. the percentage of approximated pair relations not aligned with the correspondent pair relations defined by the target rank. Of course such a measure depends on the total amount of elicitated pairwise preferences. It is formulated as percentage of the total amount of pairs without taking into account the symmetrical pair relations.

To simulate more difficult problems different target functions Φ have been configured on the same case base changing their relationships with the ranking features f_i, whose definition is strongly related to the specific case base. Problems with different complexity degree have been generated using the four categories mentioned above.

We performed many trials to assess the properties of the proposed model and overall to provide empirical evidence that it could be effective.

The aim of the first set of experiments was to assess the claim that a target order relation can be approximated taking advantage of a set of orders defined on the same collection of

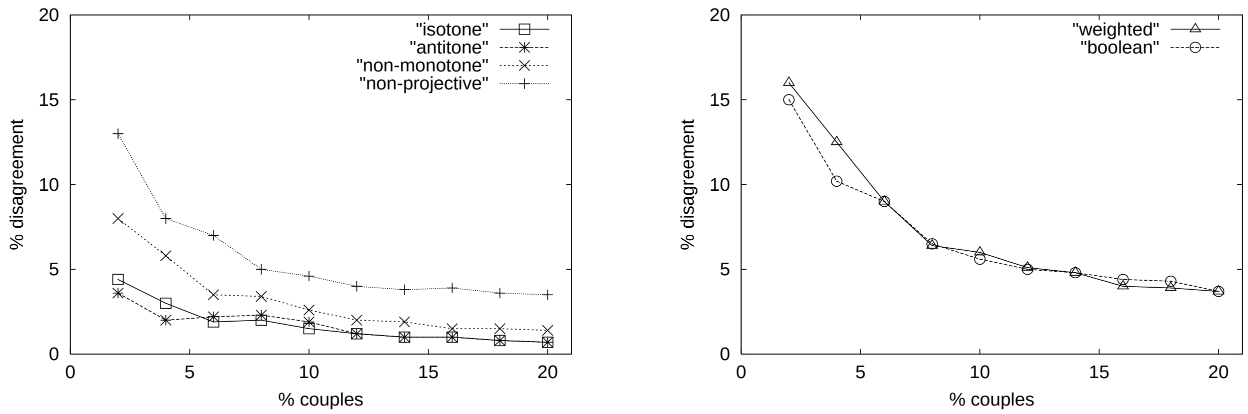

Figure 4: **Evaluation Results**: on the left side the plot of learning performances for a case base of 50 instances with different target ranks of increasing complexity; on the right side the plot of the behaviours of rated and boolean preference elicitation processes; on the y axis is shown the disagreement between the target and the approximated ranks while on the x axis is shown the amount of relations given to the learning algorithm.

cases, e.g. the ranking features. The results are illustrated on the left side of the Figure 4 where on the y axis is illustrated the disagreement measure and on the x axis the elicitation effort. The elicitation effort is formulated as the percentage of the $|C|(|C|-1)/2$ total amount of relations; the figure doesn't provide the detail of the incremental dynamic of the acquisition policy. The plots refer to the behaviour of the model on a case base of 50 instances and a case description of 20 features. The different curves correspond to different types of problem complexity without changing the case base size. The values represent the average performance over 50 trials. It is worthwhile to notice that, although it is not reported on the plots, the performance of the model is invariant with respect to the dimension of the case base and the number of ranking features.

There is a lower bound for the elicitation effort because the end users are constrained to see all the cases at least once, i.e. $|C|/2$ pairs of cases. Such a threshold is around the 2%: the leftmost point of the plots. Even though the performance is already good with the 5% of pairs, the performance of the method becomes stable around 10%, i.e. $2|C|$ pairs of cases. We can conclude that elicitating a number of preferences double respects with the number of cases it possible to lower the approximation error less than 5%. We claim that improving the acquisition policy could lower the elicitation effort achieving a number of pairs equal to the size of the case base.

The right part of the Figure 4 shows the results of the experiments related to the comparison between the boolean and the rated pairwise preferences. In these experiments we modified the configuration of the preference function Φ extending the integer values range from $\{-1, 0, 1\}$ to $[-9, 9]$. The two curves appear to be very closed to each other; this similarity increases with the cardinality of the case base, so we can conclude that the cognitive overload required by a user to provide a rated preference relation seems to be unuseful when the number of instances scale over a certain cardinality.

8 Evaluation on the Field

The off-line experiments with artificial data were aimed to assess the properties of the model and to tune the computational tools. The ultimate goal was an evaluation on the field with real users, real data and a realistic settings. For this purpose we arranged a real world experimentation in collaboration with the Geological Department of Trentino, an Italian northern region.

The experts selected the following problem: ranking a set of areas that a preliminary and rough analysis has recognized as risky from the hydrogeological point of view. The goal was to setup a plan to schedule a further detailed analysis of these areas. The challenge was to include the systemic vulnerability in deciding which areas had to be considered first. The preliminary recognition was performed using the traditional approach based on the overlay model and the result was a classification of the areas with respect to four categories: no risk, moderate risk, medium risk, high risk.

Therefore the starting point was a collection of 30 areas, mainly landslides, divided into two categories of risk: medium and high. This partition of the areas was not satisfactory to support the decision process concerning a plan for a further detailed risk analysis. The data were taken from the "Aree Vulnerate Italiane" database, a repository developed by a project led by the National Research Council of Italy.

¿From the point of view of the experimentation the goal was twofold: to confirm the partial order defined by the two categories partition and to produce a total order of the areas. For this purpose has been designed and developed a web-based application with a graphical user interface.

An area was represented by a detailed map and by a record with a quite rich description including the geographical positioning, the administrative information, the risk analysis (with data like the landslide volume, land chemical composition, the site slope, the involved infrastructures, the threatened people, ...).

As we have seen above (see Section 5) the learning process relies on a set of reference ranks. These ranks have been derived from the order relations induced from part of the features illustrated before. The rational underlying this choice is that a feature can play the role of a candidate for a simple heuristic that explains the target rank. For example if we consider the number of people involved in a natural hazard we can imagine to have a feature that encodes this information. The order relation induced by this feature could be the same of the risk value. In this case the risk assessment could be performed applying the simple rule: if the number of people involved in the area X is higher than the number of people involved in the area Y, then the area X is more risky than area Y. Of course very often it doesn't occur that a single feature explain the rationality of the risk analysis. Usually the exceptions are managed combining rules that involve the induced ranks of more features. The challenge of the learning step is to detect what are the ranking features that better explain the target rank and how to combine them together.

Experimental sessions have been performed with five experts of the Geological Department acting as a single decision maker. The typical session was characterized by the direct interaction of the users (not mediated by a technician). Although the experimentation has not been modified with respects to the definition given in the Section 4, the two steps of sampling and approximation were transparent to the users. It has been organized three sessions where the team of the expert, seated around the table, shared the output of the system projected on the wall.

The typical session was organized as follows. The users log on to the system and the initial screen proposes the agenda of the session (see Figure 5): a list of pairs, each of them composed by two alternative areas. Such a agenda is produced by the sampling policy that apply the following heuristic: the areas don't appear twice and the pairs cover all the areas. Therefore a session includes 15 comparisons. The expert were free to choose the order of pairs processing. The selection of a pair (see Figure 5) opens a new windows organized into two columns: on the left the first area and on the right the second area. Each area has three levels of details: a synthetic view (like in the snapshot), a full view, a cartographic view. An additional feature allows to enter annotation to extend or to revise the description of the area. It has to be noticed that the annotation are associated to the area and not to the pair, then when the same area will be considered for a further comparison the preference will be influenced by the annotations acquired through the previous analysis. After the comparison and the annotation the team of experts will summarize a single preference choosing one area. After the preference elicitation of all the 15 pairs the system invokes the learning algorithm to implicitly estimate the unknown pairwise preferences.

This type of session has been replicated three times for a total of 45 comparison. Sampling and learning steps were transparent to the experts while the interaction was simplified to the agenda and the comparison only.

Before to illustrate the results of the experimentation using the same measures of the off-line experiments we would like to summarize some interesting achievement. First of all the pairwise comparison allowed to assess the quality of the data and in many cases they have been revised. A second meaningful achievement is that the annotations have allowed to acquire context sensitive description of the areas. For example the presence of instruments to monitor a slope, a potential feature not included in the description of the area that doesn't apply to every types of landslide or floods. A third achievement is that the case-based analysis has allowed to detect a wrong classification: an area classified as "medium risk" has been updated to "high risk".

At the end of the three sessions we had to face with the issue of validation. Contrary to what happened in the off-line experiments here we didn't have the opportunity to know in advance the target rank and consequently to compute the disagreement with the approximated rank. For this reason we first synthesized the final rank and then we organized a fourth session to assess the coverage of the resulting order relation [Davey and H.A.Priestley, 1990]. The coverage is composed by the pairs that refer two areas strictly consecutive in the total rank. The fourth session has required to validate 29 pairwise preferences. The disagreement on the coverage of the order relation was 4%, a result obtained presenting less than 10% of pairs to the experts. These achievements, accomplished on the field in a real world setting, agree with the promising results of the previous off-line experiments.

9 Conclusions and Future Directions

In this paper we presented a case-based model for the mixed initiative elicitation process that combines a step of a pairwise case selection and a step of a pairwise preference approximation. We illustrated how a case-based approach can be sustainable also with a large number of cases whether learning phase is combined with the end user elicitation. Moreover we have provided the experimental evidence that a rankboost algorithm is effective to achieve a good trade-off between the accuracy of the final rank and the elicitation effort. A meaningful result is concerned with the boolean preference: we showed that when the number of cases increases the rated preference and the related cognitive overload become useless.

For lack of the space, we have not included a further interesting result on the characterization of the problem complexity. Four categories of increasing complexity have been introduced discriminating with respect to the relationship between the target rank and the order relations derived from the supporting funtions. A paper devoted to this aspect is in preparation.

The paper omitted illustrating the design detail of the acquisition policy. Up to now we refer to a baseline that has allowed achieving the results above. We are currently working on the enhancement of this baseline to obtain a much more effective acquisition policy that would enable a better exploitation of the available information: the by product of the boosting step and the understanding of the problem complexity.

Last but not least, after these preliminary promising results, a new challenge is open: to extend this approach dealing with the multiple viewpoints or the multiagents architecture, according to a decision theory or an artificial intellingence perspective respectively.

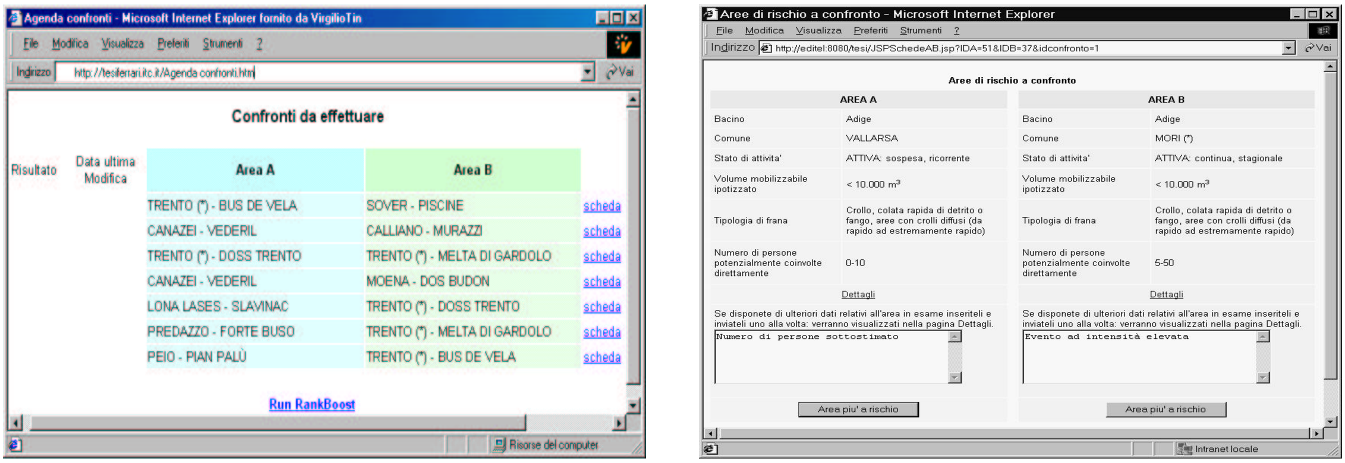

Figure 5: **The Graphical User Interface:** on the left side the agenda of pairs proposed by the sampling policy to the evaluation of the experts; on the right side the pairwise comparison to assess by contrast what area has to be considered more risky between the two alternatives.

Acknowledgments

We would like to thank the Geological Department of Provincia Autonoma di Trento and in particular Silvana Arnoldi and Roberto Cavazzana for their invaluable support to accomplish the case study.

References

[Aamodt and Plaza, 1994] A. Aamodt and E. Plaza. Case-based reasoning: Foundational issues, methodological variations, adn system approaches. *AI Communications*, 7(1):39–59, 1994.

[Beinat and Nijkamp, 1998] E. Beinat and P. Nijkamp. *Multicriteria Evaluation in Land-Use Management: Methodologies and Case Studies*. Kluwer, 1998.

[Blythe, 2002] Jim Blythe. Visual Exploration and Incremental Utitlity Elicitation. In *Proceedings of AAAI/IAAI 2002*, pages 526–532, Edmonton, Alberta, CA, July 2002.

[Caprile *et al.*, 2002] B. Caprile, C. Furlanello, and S. Merler. Highlighting hard patterns via Adaboost weights evolution. In J. Kittler and F. Roli, editors, *Multiple Classifier Systems, Lecture Notes in Computer Science 2364*, pages 72–80. Springer, 2002.

[Carrara, 1991] A. Carrara. Potentials and pitfalls of gis technology in assessing natural hazards. *Earth Surface Processes and Landforms*, 16:427–445, 1991.

[Cohen *et al.*, 1998] William W. Cohen, Robert E. Schapire, and Yoram Singer. Learning to Order Things. In Michael I. Jordan, Michael J. Kearns, and Sara A. Solla, editors, *Advances in Neural Information Processing Systems*, volume 10. The MIT Press, 1998.

[Davey and H.A.Priestley, 1990] B.A. Davey and H.A.Priestley. *Introduction to lattices and order*. Cambridge University Press, 1990.

[Devetag and Warglien, 2002] G. Devetag and M. Warglien. Representing others' preferences in mixed motive games: Was schelling right? *to appear*, 2002.

[E. J. Henley, 1992] H. Kumamoto E. J. Henley. *Probabilistic Risk Assessment*. IEEE Press, New York, 1992.

[Freund and Schapire, 1999] Y. Freund and R. Schapire. A Short Introduction to Boosting, 1999.

[Freund *et al.*, 1998] Y. Freund, R. Iyer, R. Schapire, and Y. Singer. An Efficient Boosting Algorithm for Combining Preferences. In *Proceedings 15th International Conference on Machine Learning*, 1998.

[Ha and Haddawy, 1998] Vu Ha and Peter Haddawy. Toward case-based preference elicitation: Similarity measures on preference structures. In Gregory F. Cooper and Serafín Moral, editors, *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI-98)*, pages 193–201, San Francisco, July 24–26 1998. Morgan Kaufmann.

[Linden *et al.*, 1994] G. Linden, S. Hanks, and N. Lesh. Interactive assessment of user preference models. In *User Modeling*, 1994.

[Roberts, 1979] F. Roberts. *Measurement Theory with Application to Decision Making, Utility and Social Sciences*. Addison Wesley, Reading, MA, 1979.

[Saaty and Vargas, 1994] Thomas L. Saaty and Luis G. Vargas. *Decision Making in Economic, Political, Social and Technological Environments With the Analytic Hierarchy Process*. RWS Publications, 1994.

[Saaty and Vargas, 2000] Thomas L. Saaty and Luis G. Vargas. *Models, Methods, Concepts & Applications of the Analytic Hierarchy Process*. Kluwer Academic, 2000.

[Saaty, 1999] Thomas L. Saaty. Fundamentals of the analytic network process. In *Proceedings of International Symposium on Analytical Hierarchy Process*, 1999.