A methodology for the development of a Knowledge Based Judging System for free response Materials

Dick J. Bierman
University of Amsterdam

Abstract

A methodology is proposed to elicit (intuitive) knowledge used by expert-judges of free response materials. This methodology is an extension of the knowledge elicitation procedure which has been developed for classification (1) tasks like Psychodiagnostic classification. The normal judging procedure, which is a matching task (2), is modelled as a double classification task. Each of the classifications (of the target as well as of members of the target-set) is moderated by deep knowledge which accounts for the interpretation of interacting elements in the material. This latter is the knowledge that transcends the superficial visual correspondence like symbolic meaning. The knowledge elicitation method will be based upon the presentation of the judgements of old cases to two distinct learning systems. This results in the elicitation of global knowledge and of the detection of pathological cases (3) in which the deep knowledge is hidden.

The matching task, modelled as a 'moderated' double classification, has a different flow of control from a straight classification task. The consequences for the resulting expert-system's control structure are discussed too.

---

(1) A (double) Classification task is a task in which instances have to be classified in one of a limited set of categories.

(2) A Matching task is a task in which two instances from a set of instances have to paired.

(3) A pathological case, in the present context, is a case where the Knowledge Based systems judgement differs considerably from the human expert's judgement.

1. Introduction

It has been found that certain judges perform consistently better than others while matching targets with a target-set (Schmeidler, private communication). It seems unlikely that this is purely due to psi of the judge since psi generally does not display consistent behavior. Therefore it might be hypothesized that it is the (intuitive) knowledge of the specific judge which accounts for his/her better performance on this task. It has been proposed (Morris, 1986) that the use of expert-systems might help psi-researchers in tasks where they lack in expertise like in the detection of fraud. He argues that the expertise of macigians could be formalized in such a system and be made available to each individual researcher. Similarly the expertise of the best judges of free response material could become available through implementation of a knowledge based free response judging system. This use of techniques from the field of Artificial Intelligence to represent scarce knowledge, should not be confused with the use of AI-techniques for the representation of free-response material (Maren, 1986). Although it is not explicitly mentioned by Maren, the reference that she makes to machine vision, strongly suggests that the matching of the protocol and the target is mainly seen as a **visual** process. According to Maren the free response material should be represented in the form of trees in which the nodes are perceivable 'objects' like 'flames' and the links represent relations like 'adjacent to'. The matching is proposed as a form of tree-matching. Apart from the fact that it is well known in the field of AI that this type of matching is rather unreliable, it seems to me that it is also rather superficial. We expect that focussing our attention on the (knowledge used in the) human matching process might reveal the more fundamental information about the role of the **meaning** of the material. It is striking that in Maren's proposed representation of the complex target material only visual features are present. Actually the type of visual matching that Maren proposes to be done by a machine can be better performed by **any** non-blind human.

It should be remarked however that the crucial element in the development of expert-systems is **not** the implementation of the system but the elicitation of the knowledge that has to be plugged into the system . In the case of knowledge about trickery for instance it is doubtful if one can find experts who are willing to transfer their knowledge. A part from that, the detection of trickery is largely driven by visual information. The proper representation of this visual knowledge might also be a major problem in this domain of expertise. In the case of free response judging one can expect cooperation of the expert-judges. Although the material is also visual there are strong indications that simple keywords are able to represent these pictures satisfactorily. This conclusion can be drawn from the analytical judging procedures developed by Jahn et al (Jahn et al, 1980).

2. Analytical judging versus Knowledge Based judging

It has been found that simple (linear) regression formula make predictions comparable or better that human experts in the domain of Psychodiagnostics (eg. Schmidt & Hunter, 1981). Thus it is not surprising that the analytical judging procedure which is very similar to a approach by linear regression does also yield satisfactory results. However it should be noted that although the average performance is adequate this approach fails in pathological cases. It appears that this is due to the failure to take into account any interaction between the predictor variables. In the analytical judging procedure for instance the simultaneous occurrence of two elements is counted as the sum of the scores for the case that they occur alone. Thus if two elements together would have a symbolic meaning while on itself they don't have such meaning this is missed in the analytical judging procedure. A knowledge based judging system is capable to represent and use this type of knowledge.

3. Matching as Classification task

It is found that most problem solving tasks can be seen as classification tasks. For instance the problem to determine if a chess end game is 'winning' or 'undecided' or 'loosing' has been treated as such. Even activities such as robot-planning can be recast as a classification problem (Dechter & Michie, 1985). In the case of matching of free response material from psi-experiments however there is a special problem. Since the categories 'correct-match' and 'incorrect-match' in psi research are solely determined by chance, these categories do not have objective attributes. Thus the task can not be modelled as a direct classification task. Therefore we propose to model the matching process as a double classification process.

The judge is thought to begin with a classification of the protocol into one of his internalized categories. Secondly this procedure is repeated for each of the members of the target-set. Finally the results of these classifications are evaluated using overlap-measures. If no clear-cut match can be made a secondary evaluation is done which takes into account (subtle) interactions between attributes.

4. Knowledge Elicitation methods

It is generally acknowledged that the elicitation of knowledge which is needed to drive expert systems is a 'bottleneck problem' (Feigenbaum, 1981). The traditional interview approach is said to yield only a few rules per man-day (Quinlan,1986). Although this figure applies to the In-depth knowledge which is elicited near the end of the procedure, this was reason enough to stimulate the research in machine learning methods as a means of explicating knowledge (Michie, 1983). Very often this rather unstructured approach is accompanied by so-called rapid prototyping. This means that the system is implemented while the knowledge base is essentially of low quality and incomplete. Wielinga and Breuker (1984) have argued that this might results in poor final systems like most rule-based systems to date. If this rapid prototyping results in poor systems for rather well understood areas of human expertise it seems unwise to use it for the task of free response judging. Wielinga and Breuker have proposed a more structured methodology in order to make it more transparent and thereby more efficient. In this approach emphasis is given to the necessity of a well specified framework for interpretation of the verbal material, be it interviews with, or thinking aloud protocols produced by, the expert. In the present paper it is proposed to combine the structured knowledge elicitation procedure with the use of learning systems.

5. Proposed Procedure

The proposed methodology differs from the accepted methodologies of knowledge elicitation by using information already present in the database of previously classified cases. The elicitation procedure consists of three major parts:

1: Learn
2: Pathology-detection
3: Confrontation

In the first phase the expert-judge will be interviewed on the set of attributes which are used to describe a target-picture. Also the primary set of classes is formulated. After that a training set of old cases is selected to be presented to a learning system. Each case consists of a series of attribute values together with the classification by the expert-judge. After the training, the systems are able to classify other cases from the old database and to compare classifications of the target-set with the classification of the protocol. The trained system has become a (first order) model of the Expert-judge.
In the second phase the remainder of the old database is presented to the 'trained' system for judging. If the judging by the system differs from that made in the past by the human Expert we call this a 'pathological case'.

In the third phase the human Expert is confronted with the set of pathologies. The knowledge Engineer might directly ask the Expert why he deviated from the model or give him the cases to solve again while thinking aloud. Analysis of the thinking aloud protocol should occur in terms of deviations from the model and thus produce additions to the knowledge base.

A schematic comparison between the traditional methodology and the proposed methodology is given in fig.1.

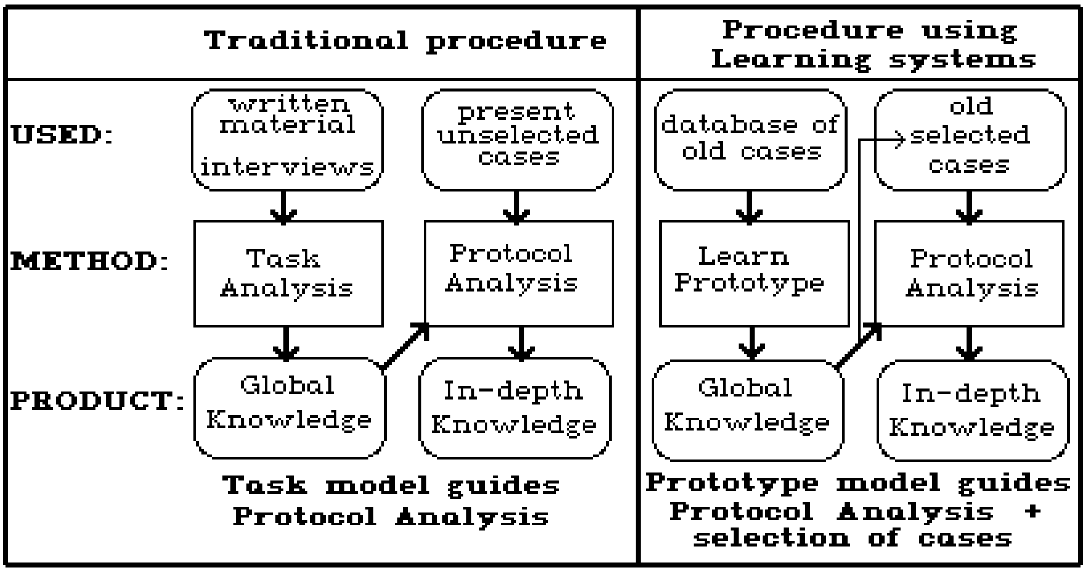

Figure. 1

6. The learning systems

Previous work which tried to apply learning systems in the process of knowledge acquisition used systems like ACLS (Automated Concept Learning System), which construct a decision tree from examples. However it was found that, although the resulting decision trees were able to classify new cases properly, these trees, which represent the knowledge of the human expert, very often were hardly recognized by the same expert. This decision tree representation offered therefore not a fruitful framework for the Knowledge Engineer to base his further interviews on. This situation is not very different from a representation by linear regression models which have shown to have considerable predictive power (eg. Schmidt & Hunter, 1981). However the linear regression formula does not make a lot of sense to the human expert. Therefore it has been proposed (Bierman and Akkerman, 1986) not only to use a ACLS type of learning system but also to use a learning system that is supposed to create a psychological valid representation of the human expert's knowledge.

6.1 The Prototype-learner

The 'prototype' model has been developed by Rosch (Rosch, 1978). In contrast with linear regression models, the 'prototype' model allows for non-monotonic relations between the values of the attributes and the class-determination. A system has been implemented that is capable of learning categories as proposed in the Rosch model. During the learn-phase a training-set of old cases consisting of the values of the attributes and the resulting classification are offered to the system. The system learns which attributes contribute to which degree to the final classification decision. After the learning phase new cases can be offered to the system which will calculate a overlap-score of the new instance with the 'prototype' of a class.

6.2 The decision-tree learner

A decision-tree learning system similar to ACLS has been implemented. The choice of the root-attribute (and subsequent layers) which is one of the crucial operations in this type of models, is done on the basis of maximizing informational contents (through a chi-2 test for stochastic independence). If all attributes have the same number of potential values, this is an appropriate choice (Kononkenko et al, 1984). We will not use any of the mechanism which are thought to prevent noise from entering the data since one of our goals is to detect 'pathological' cases. The distinction between noise and a pathological cases can not be made by any of these noise reduction mechanisms.

7. Concluding remarks

Current work using a similar knowledge elicitation approach in the domain of Psychodiagnostics is promising (Bierman, 1987). In this research the knowledge that a selection Psychologist uses to classify a candidate as suitable (or not) is elicited . This classification task is supposed to have clinical aspects and some Psychologists claim that they use their intuition during this task. It turns out that the use of two learning systems results in converging evidence with regard to which cases are to be considered as pathological and therefore deserve further attention. If a case which is considered by the prototype based judging as a pathological case turns out to have extremely long branches in the decision tree, this is an extra indication that there is something special about that case. Actually the long branch contains information about the interactions which might play a role for this particular case and this information might be used to interpret the thinking aloud protocol of the expert-judge. It appears that 'intuitive' knowledge can be elicited and implemented as a moderator of a primarily pattern-matching based classification.
A final remark should be made with regard to the control structure of the knowledge based judging system. In most interesting cases there will be no clear-cut matching on the basis of the prototypical classes. Thus control should be passed to a secondary evaluation based upon complex production rules. The interaction between those two evaluations is still a matter of research. None of the commercially available expert-system shells allow for such interplay between two evaluation mechanisms. Therefore it can not be expected that a serious knowledge based judging system will be available before the 90'ies.

References

Bierman, D.J. (1987). The use of learning systems as a tool for the elicitation of Knowledge. *Submitted for presentation at IJCAI-87*, Milano.

Bierman, D.J & Akkerman, A.E. (1986). The development of Knowledge Based Psycho Diagnostic Systems. In: *Proc. of the 2nd Workshop on Computers in Psychology.* May 1986, Nijmegen, The Netherlands. Swets & Zeitlinger (Lisse), In Press.

Dechter, R. & Michie, D. (1985). Sructured induction of plans and programs (Technical Report). IBM Scientific Center, Los Angeles, CA.

Feigenbaum, E.A. (1981). Expert systems in the 1980s. In A.Bond (Ed.), *State of the art report on machine intelligenc.* Maidenhead: Pergamon-Infotech.

Jahn, R.G., Dunne, B.J. & Jahn, E.G. (1980). Analytical judging procedure for remote perception experiments. *J. of Parapsychology,* **44,**207-232.

Kononkenko, I., Bratko, I., & Roskar, E. (1984). Experiments in automated learning of medical diagnostic rules (Technical Report). Jozef Stephan Institute, Ljubljana, Yugoslavia.

Maren, A.J. (1986). Representation and Performance Evaluation Approaches in Psi Free-response Tasks. Presented at the PA-86 convention.

Michie, D. (1983). Inductive rule generation in the context of the Fifth Generation. *Proc. of the 2nd Int. Machine learning workshop.* Univ. of Illinois at Urbana-Champaign.

Morris, R (1986). Minimizing Subject Fraud in the Laboratories. *EJP, V6-2.* p.137-151.

Quinlan, J.R. (1986). Induction of Decision Trees. *Macine Learning 1:* p.81-106.

Rosch, E. (1978). Principles of Categorization. In: E.Rosch & B.B. Lloyd (Eds.): *Cognition and Categorization,* Hillsdale N.J.: Lawrence Erlbaum, 1978.

Schmidt, F.L., Hunter, J.E. (1981). Employment Testing, old theories and new research findings. *American Psychologist, Vol.36-10,*p.1128-1137.

Wielinga,B.J. & Breuker, J. (1984). Interpretation of verbal data for knowledge acquisition. *Proc. of ECAI-84,* 1984, p.41-50.