EXPERTISE: AN OPTIMIZED DATA MODEL?

Heikki Hyötyniemi

Helsinki University of Technology, Control Engineering Laboratory
Konemiehentie 2, FIN-02150 Espoo, Finland

This paper presents a new approach to case-based reasoning and knowledge representation. The scheme is based on special assumptions about the data ontology: how the information structures are manifested in the observation data. It turns out that some of the old paradoxes can be attacked in this framework, like the shift from novice to expert. As an example, an industrial instrument selection database is modeled along the presented guidelines. The low-dimensional data model makes it easier to evaluate the consistency of the knowledge base, and it also facilitates automatic rule revision.

1. INTRODUCTION

1.1 Dilemmas of human expertise

The knowledge representation is normally implemented using rule-based formalisms. Of course, this is a compact way to express knowledge in rather limited domains, when the rules can be explicitly defined. However, using the traditional AI approaches, many of the observed features concerning human problem solving cannot be explained.

It is a well-known fact, for example, that the experts have not stored their knowledge in the linguistic or rule-based form — and it is also a well-known fact that often this expertise cannot be expressed explicitly at all. Another mystery is how the shift from novice to expert could be explained: even though the expertise cumulates, the expert responses seem to become faster and more and more mechanized as compared to the novice (see [1]). Rule systems become clumsier and slower if new rules are added.

Yet another problem in the traditional knowledge formalisms is the "crisp" nature of the representations. The binary dependencies are rigid, they cannot represent the fuzziness of categories, and their performance degrades abruptly rather than gracefully. It is no wonder that the new formalisms like fuzzy systems and probabilistic networks have lately been studied extensively. Another very different approach to knowledge representation and application is offered by *Case-Based Reasoning.*

1.2 Case-Based Reasoning

Case-Based Reasoning (CBR) is an implementation of the idea of "solving by analogy" [5]. Past exemplars are stored as cases in a knowledge base, and they are used as frames for solving new, more or less analogical cases. CBR is partly founded on psychological plausibility theory: the role of specific, previously experienced situations in human problem solving has been demonstrated (for example, see [8]).

The idea of CBR is appealing because the knowledge representation is extremely transparent, and, in principle, the reasoning process using the cases is simple. Whereas the complex rule systems are plagued by the surprises caused by nonmonotonicity of the reasoning (ordering or other slight modifications in the rule base can result in very much differing outcomes), the associative search in CBR is rather robust.

There are many implementations of CBR; because of the wide variety of applications, it seems that it is only the core concepts that remain invariant. The intuitively appealing ideas behind CBR have been studied in various other areas, too — for example, there is some connection to the ideas of "semantic maps" presented in [7]: semantic concepts are described in terms of a few attributes, and using these attributes as input data, automatic classification in categories is carried out using a self-organizing map.

When seen from the point of view of cognitive science, it seems that some of the dilemmas have been relaxed:

- Reasoning in CBR is associative, non-verbal, and there is no need to express knowledge in the explicit rule-form.
- In a sense CBR implements the category fuzziness; when an appropriate metric is defined, there are methods to measure how "near" each other different cases are.
- However, the question about the shift from novice to expert is still left open: as new cases are stored separately, the search process becomes slower and slower, contrary to human expert behavior.

The basic problem in CBR is how to define the "case space" and how to define a reasonable metric in a space of symbolic attributes. How the more or less semantic attributes can be compared against each other? When are two cases "near" each other, and when are they far apart? It turns out that addressing this fundamental question may also help in attacking the "novice vs. expert" challenge.

1.3 Models for expertise

The hypothesis here is that the human experts, starting also from "tabula rasa", have seen many instances of cases, and their internal model of the problem field has adapted to match the properties of the observation data distribution. The goal boils down to *finding an model for the problem domain.* This idea needs to be emphasized:

> Rather than constructing models for the mind, now *mind is seen as a model.*

The role of a model is to abstract and compress the data, finding underlying redundancies or relationships between data entities. It is now assumed that the difference between an expert and a novice is that the expert has an *optimized* model of the environment. This modeling view is motivated in [2] and [3].

If it is only question of model optimality, the problem boils down to first finding an appropriate model structure and thereafter the model parameters. The nice thing about this is that the well-founded methodology of systems analysis and modeling are available to us[1]. Compared to simpler data modeling tasks, the complexity is now caused by the high dimensionality and large number of modalities that need to be mastered in the same framework.

[1] There are various views of what a *model* is — for example, the "mental models" of cognitive science are far from the systems theoretical modeling, where everything is based on explicit optimality criteria and concrete mathematical optimization schemes for adapting the models.

2. MODELING COMPLEX OBSERVATIONS

2.1 Data ontology

The assumption about the domain area data distribution can be expressed in statistical terms: it is now assumed that the variation in the data can best be captured using a combination of *clusters* and *independent components.* The clusters define the coarse structure, whereas the independent components span the axes of internal fine structure within the clusters. Interpreting this in cognitive terminology, it is assumed that hidden parameters, or features, modify the category prototypes towards the actually observed individual. There is some evidence that the unexplicable expert knowledge really could be based on this kind of mental structures (see 4).

The optimized feature vectors implement "atoms of excellence", defining the directions of "admissible variations", thus redefining the distance metrics. As compared to the basic CBR scheme, this fine structure extension enhances the expressive power essentially. It seems that usually it is the features that reveal the most interesting hidden structure underlying the data samples.

The algorithm for implementating the above view (and also some examples) is presented in [2]. The idea is to extract the correlation structures between the input variables assuming they are *sparse coded;* if the ontological assumption holds, these "dependency dimensions" θ_i can be utilized as features for reconstructing the original observation vector f:

$$\hat{f} = \theta_1 \phi_1(f) + \cdots + \theta_N \phi_N(f). \tag{1}$$

In practice, some of the features (the most significant ones) determine the cluster or category center, whereas the minor features only modify it (their weights being determined by ϕ_i). The features that are not utilized have zero weight.

2.2 Model interpretation

If the expert knowledge is stored as a "mental image" that has the presented structure, finding the missing pieces of information from it can readily be accomplished using associative search strategies based on fitting against the model. Inference can be interpreted as the problem of finding a point in a high-dimensional space when only a subset of variables are known beforehand. Given an incomplete data vector the matching process reconstructs the "most plausible" values for the unknown quantities as predicted by the model.

The shift from novice to expert can be nicely explained in this framework. In the beginning, when no specialized data structures exist, declarative knowledge consisting of example cases are directly stored as features of their own. Using these example cases, the construction of the "mental view" is carried out as in a production system, one associative inference step at a time[2]. The examples constitute the set of compact, non-optimized features. Later when the features become optimized and the correlation structures between data items are exploited, various inference steps are no more needed; however, the reasoning mechanisms in both cases are still the same.

[2]For example, if there are the examples "Tweety is a bird" and "bird sings", starting from "Tweety" it takes two steps to find out that Tweety sings

3. EXAMPLE APPLICATION

3.1 Selecting a valve

Valves are among the most common devices in process industry. There exists a special database[3] containing knowledge of how to select a valve for different pulp and paper industry applications. There is information about the operating environments, like flow temperatures, pressures, densities, etc.; along with this data there is information about the appropriate valve types, materials, and size specifications. All rules are of the same form, even if only a subset of the attributes are normally employed. Over 600 prototypical selections have been stored in this database; this knowledge has been collected by different experts during many years, and there is no single person who would now have a thorough understanding of this information. There may be rules that are outdated or contradictory, but how to spot them? A consistent approach to maintaining the acquired information would be invaluable.

A totally autonomous loop from observations to knowledge base adaptation cannot be realized: a human is needed to analyze the causal dependencies and verify the actions, and the human has to take the final responsibility. But a "smart" system could preprocess the data and deliver manageable chunks of information to the human expert, perhaps helping in the process of gaining intuition and understanding. The proposed knowledge modeling scheme might be step in that direction.

3.2 Coding of information

Because the rules always employ the same variables (if the "don't care" values are ignored) they can easily be represented in a homogeneous case format. In this example, the rules are coded in a straightforward manner: all numerical entities are included in the "case vector" as such (scaled so that the maximum value in the database becomes unity), whereas there are various entries allocated for each variable containing qualitative information (these entries in the data vector are binary: "1" meaning that the variable has the corresponding value, and "0" otherwise). This way the data vectors become 86 dimensional.

Figure 1 visualizes the mnemonic names for data sections are defined (for example, "Press" stands for the pressure level, and "Mat$i$" stand for the materials used in different parts of the valve). In all of the subsequent figures the interpretations of the vector elements remains the same. The first 7 entries in the vectors are numeric, containing information about pressure, temperature, etc. The next 13 entries (from 8 to 20, denoted "Mode") contain *a priori* rule classification codes, and the rest of the entries concern the valve properties.

When the algorithm is applied using the set of case vectors as input, results are dependent of the weighting of the vector elements. As no *a priori* understanding is available about the relevance of the attributes, each variable has now equal weight.

3.3 Results

Three clusters and three fine-tuning features were extracted from the data; two of the three features were used simultaneously. The resulting cluster centers and the fine tuning features are shown in Figs. 2 and 3, respectively.

When trying to gain intuition from the extracted model, it is the features that are

[3] Property of Neles Automation

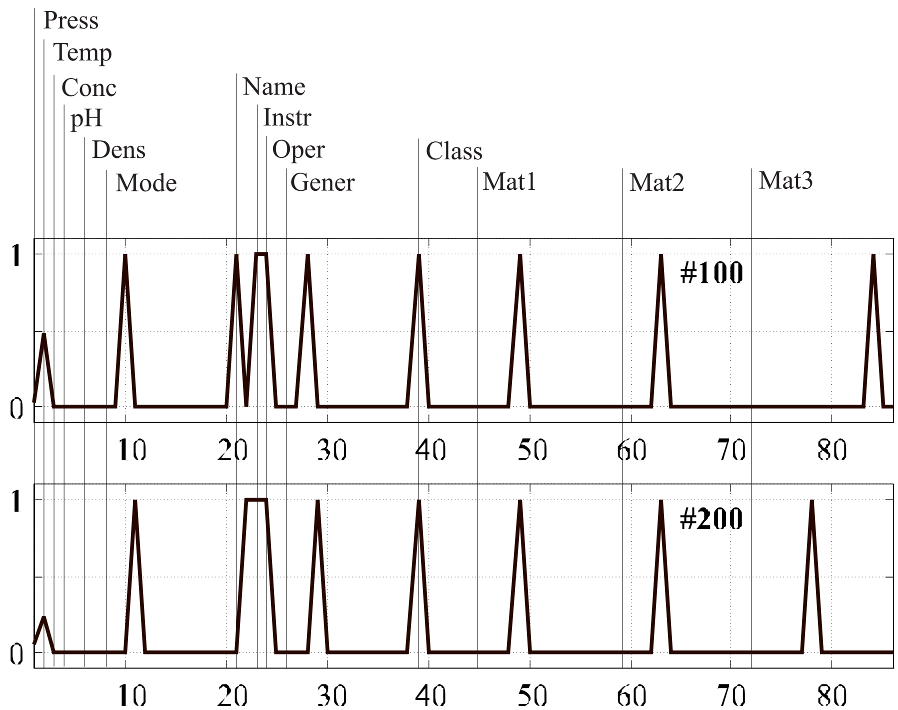

Figure 1: The interpretation of the "rule vectors", and two sample rules coded (rule indices #100 and #200)

usually the most important structural components. For example, let us study feature #2 little closer: it seems that the entries that have the most emphasis are in the "Oper" section — these variables reveal whether the valve is used in manual mode (entry #24 having value "1") or as a control valve (entry #25 being "1"). These two options are mutually exclusive (this is reflected in the opposite weights), and the typical differences between these two operating modes are recorded in this feature vector. For example, it seems that control valves are typically used when concentrations are relatively high (entry #3 having positive correlation with #25). Feature #3, on the other hand, can be interpreted as representing the changes in the valve materials "Mat1", "Mat2", and "Mat3" as a special case of "Mode" is selected (entry #11).

In Fig. 4, the model is being tested: how well the extracted six vectors (only two of which can be simultaneously used) can capture the knowledge. Below is a table corresponding to the figure: the coordinates of the selected rules are shown. The reconstruction is then created as a weighted sum of the prototype vectors, as defined in (1).

	Data vector index					
	100	**200**	**300**	**400**	**500**	**600**
Cluster 1	0	0	1	1	0	0
Cluster 2	1	0	0	0	1	0
Cluster 3	0	1	0	0	0	1
Feature 1	0.59	0.57	0	0	-0.81	0.49
Feature 2	0	0	-0.29	0.97	0	0
Feature 3	0	0	0	0	0	0

The results seem to be rather promising: it seems that the set of over 600 rules can to some extent be captured in a model of only three cluster centers and three additional

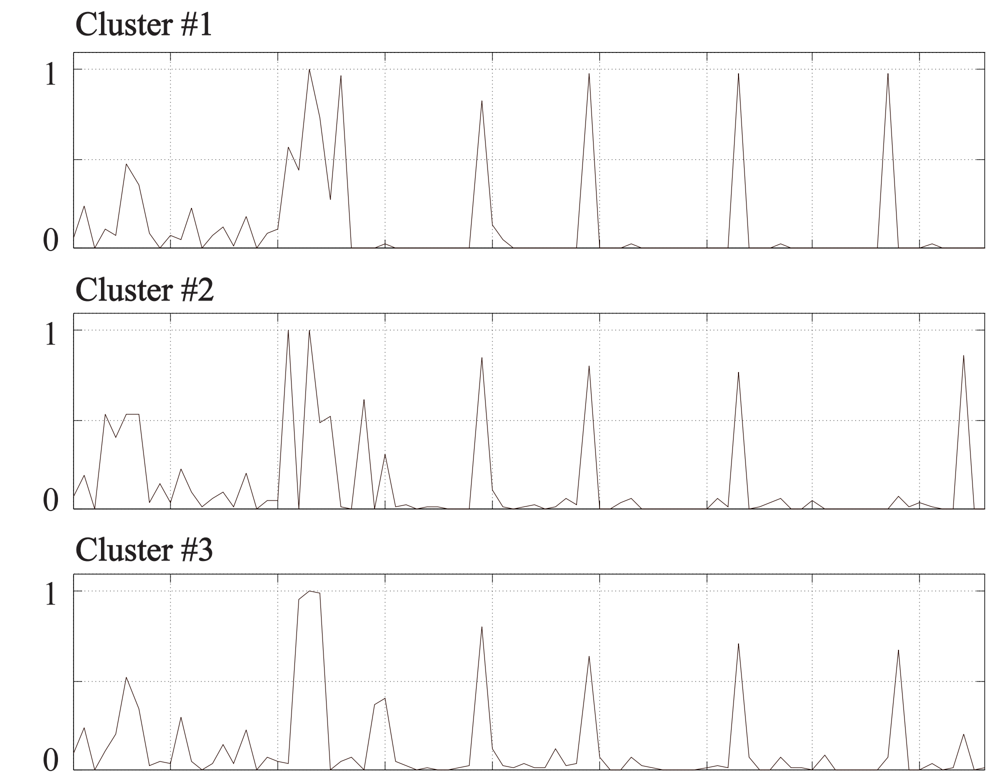

Figure 2: The three clusters (or "categories") that are extracted. These are the prototypical centers that are modified using features for "fine tuning" (see Fig. 3)

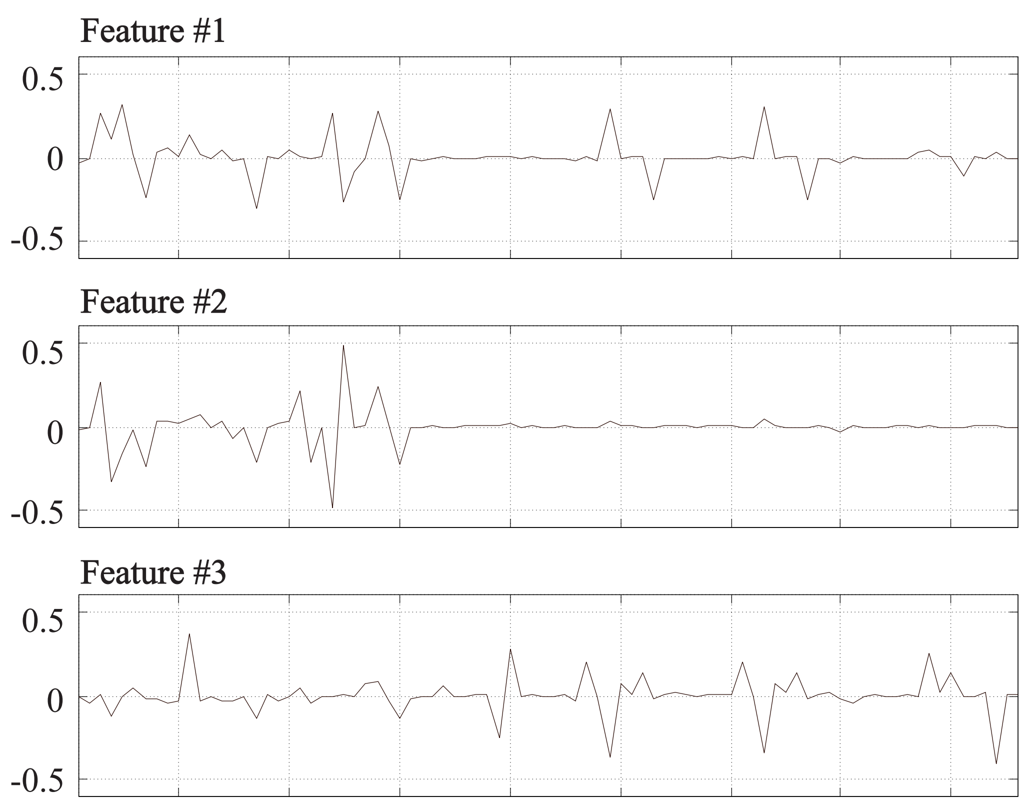

Figure 3: The three fine-tuning features. It is interesting to note their mutual independence and different roles in explaining the data

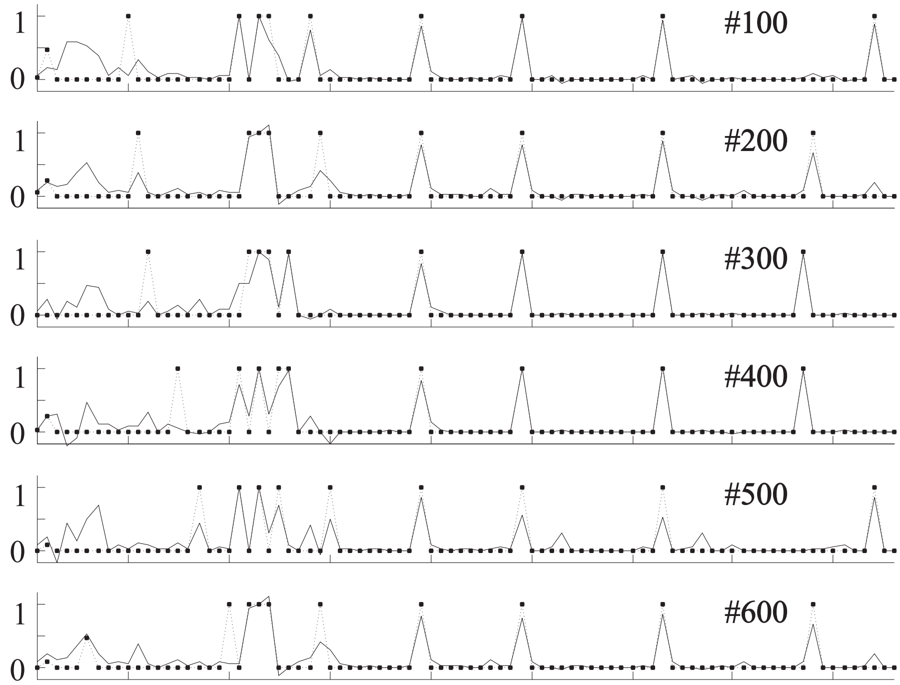

Figure 4: Six randomly selected rules (the rules #100, #200, #300, #400, #500, and #600, respectively) are shown (dotted line), and their "reconstruction" (solid line) using the created model

features. This kind of very compressed view into the valve selection expertise helps in gaining intuition and check the consistency of individual selection rules. As compared to standard CBR, only three "cases" are now stored (the cluster centers); the additive features enhance the expressive power of the model remarkably (assuming, of course, that there really exists some hidden structure that can be revealed).

Some of the vector entries cannot be reconstructed — specially the entries 8 – 20, meaning that the *a priori* "Mode" classification is perhaps not justified. Note that the discrepancies in the entry region 1 – 7 do not necessarily indicate model mismatch; these values in the rules are not always given (assuming some default values), and the missing values are shown in the figures as zeros. If there are multiple peaks in the region 21 – 86, or if the peaks do not extend to value 1, this means that there is (according to the model) some evidence that the valve selection is not absolutely certain.

3.4 Further developments

Whereas the traditional rule-based formalisms do not allow flexible adaptation of the rule bases, the presented "continuous-valued" approach to knowledge representation offers nice new possibilities to utilizing feedback from the field.

One could, for example, incorporate an additional input, a "performance index", in the augmented input data (assuming that this kind of responses were available). This performance index would contain the information about the practical experiences concerning that specific valve, like the maintenance need and failure rate, and this performance rating would be adapted as experiences are gathered. The model with this additional information could be used, for example, to carry out performance prediction ("what if" analyses).

From the point of view of reasoning based on existing cases, interesting new problems arise. For example, how to ensure “rich” valve performance information? There are many valves in the field, all delivering information, sure, but all of the valves have been selected using the same rules — the rules to be updated. Using the terminology from identification theory, this can be compared to a case where one is trying to identify a system in a closed loop. There is not enough excitation in the observations if the data is collected from plants that have identical instrumentation. Some kind of “controlled variation” in the valve selection is perhaps needed to supply for useful information; this is actually an *experiment design* problem.

REFERENCES

1. Elio, R. and Scharf, P.B.: Modeling novice-to-expert shifts in problem-solving strategy and knowledge organization. *Cognitive Science,* **14**, 1990, pp. 579–639.

2. Hyötyniemi, H.: On the Statistical Nature of Complex Data. In *SCAI'97 — Sixth Scandinavian Conference on Artificial Intelligence; Research Announcements* (ed. Grahne, G.), Helsinki University, Department of Computer Science, Report C-1997-49, Helsinki, Finland, 1997, pp. 13 – 27.
   Available from `http://saato014.hut.fi/Hyotyniemi/publications`.

3. Hyötyniemi, H.: On Mental Images and 'Computational Semantics'. In *Proceedings of the 8th Finnish Artificial Intelligence Conference STeP'98* (eds. Koikkalainen, P. and Puuronen, S.), Finnish Artificial Intelligence Society, Helsinki, Finland, 1998, pp. 199–208.
   Available from `http://saato014.hut.fi/Hyotyniemi/publications`.

4. Hyötyniemi, H. and Hasu, V.: Independent Components of Expertise. In *Proceedings of the 9th Finnish Artificial Intelligence Conference STeP 2000,* Espoo, Finland, August 29–30, 2000.

5. Leake, D.B. (ed.): *Case-Based Reasoning: Experiences, Lessons, and Future Directions.* MIT Press, Cambridge, Massachusetts, 1996.

6. Pearl, J.: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* Morgan Kaufmann Publishers, San Mateo, California, 1988.

7. Ritter, H. and Kohonen, T.: Self-organizing semantic maps. *Biological Cybernetics,* **61**, 1989, pp. 241–254.

8. Ross, B.H.: Some Psychological Results on Case-Based Reasoning. *Proceedings of the Case-Based Reasoning Workshop,* Morgan Kaufmann Publishers, 1989, pp. 144–147.