Qualitative Models as Indices for Memory-based Prediction

Boi Faltings

Artificial Intelligence Laboratory (LIA)

Swiss Federal Institute of Technology (EPFL)

IN-Ecublens, 1015 Lausanne

Switzerland

faltings@lia.di.epfl.ch

October 31, 1996

1 Introduction

Computational modelling has now reached a level of maturity where it is possible to correctly simulate the behavior of a new airplane design and thus dispense with building prototypes. However, many processes are not understood enough to construct such detailed models. Well-known examples of this are cement kilns or the combustion processes in an automobile engine. It is clear that enormous benefits can be reaped if it were possible to correctly model such processes.

The models which allow us to simulate airplanes are the result of many years of scientific work based on careful observations of many *examples*. Experimentation allowed choosing examples to prove or disprove specific hypotheses. This made the process of learning models from the observations computationally tractable.

For many of the systems which we do not know how to model, it is difficult to construct experiments giving just the crucial conclusions. Thus, even though observations might contain all the information necessary to learn a
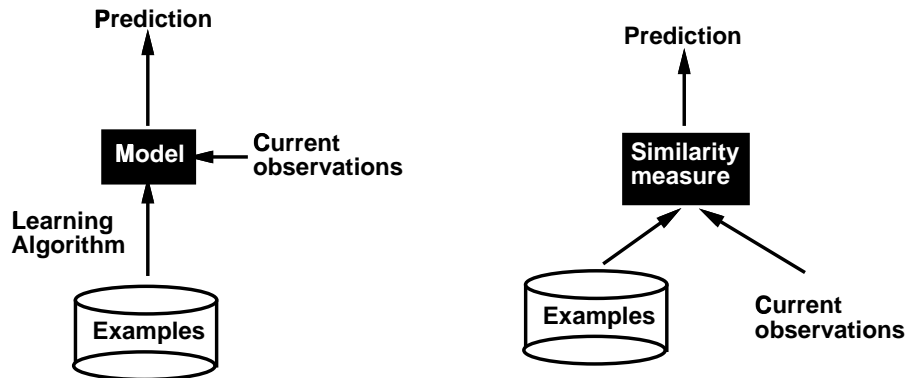
Figure 1: *Modelling by induction, and modelling with memory-based reasoning.*

correct model, it is not possible to choose the examples in a way which makes this learning process computationally tractable and error-free.

The question then arises whether it is even necessary to learn models at a great level of precision. The inductive learning problem in general is of exponential complexity, so for complex systems any learning algorithm can only provide approximations whose value is uncertain.

It has been observed that human experts often reason from particular previous experiences. An abundant source of examples for this can be found in the financial sections of newspapers, where the performance of certain investments is habitually explained in terms of similar cases in earlier history. Note that this elegantly avoids the combinatorial complexity of inductive learning: the time required for finding a similar example grows at most linearly with the number of examples. Even though this retrieval operation has to be performed once for each new situation, and thus on the order of once for each example, the resulting complexity is at most the square of the number of examples which is still much better than any inductive learning scheme. The paradigm of *memory-based reasoning* ([3]), sometimes referred to as *instance-based learning* ([1]), is intended to replicate such reasoning on a computer (Figure 1). The difficulty now shifts from modelling the relation between observations and a value to be predicted to deciding when a previous example is similar enough to make a good predictor. This question, intuitively simple, turns out to be very subtle and sensitive to a careful cho-

ice of attributes and their weighting. Similarity measures which fully exploit the information in the examples require a lot of knowledge about the system which generated them!

Luckily, in most cases such knowledge is in fact available in the form of qualitative models. At the qualitative level, most systems are quite well-understood: we know that the more fuel we burn, the higher the force generated by the combustion. Even systems which are termed "chaotic", such as the weather, are actually quite stable and predictable at a sufficiently high qualitative level. Thus, the question becomes how to exploit qualitative models in similarity metrics for memory-based reasoning. This is the topic of this paper.

I first briefly discuss the strengths and weaknesses of the qualitative and memory-based reasoning paradigms. I then show how the two can compensate each other's weaknesses. We have applied this idea to two different processes: coffee roasting and decaffination. Both applications used large amount of data collected from actual plants operated by Nestle in the UK and Spain. They are examples of processes for which existing models can provide only very inaccurate predictions. In both cases, attempts to predict behavior using statistical methods and neural networks have not succeeded in providing usable predictions. In contrast, the qualitative models used in memory-based reasoning take into account subtleties of the process which purely statistical criteria are likely to miss. The results are thus significantly better than what could be obtained with conventional methods.

2 Qualitative Reasoning: structured but imprecise

Qualitative Reasoning attempts to model the way that human experts make predictions in continuous systems without precise numerical information. This involves two parts: a qualitative *model*, and a *reasoning* technique for generating predictions.

Qualitative models have so far been investigated almost exclusively for systems with *lumped parameters*. Formalisms for qualitative modeling ([4, 11, 6]) differ in many aspects but agree on the following elements:

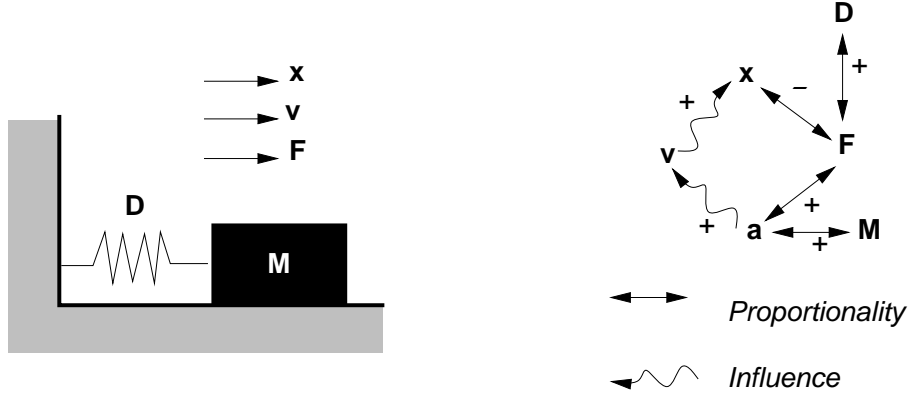- *parameters* which take on *qualitative values*, for example positive (+),

3

Figure 2: *A mass on a spring, and its qualitative model.*

zero (0) and negative (-).

- *qualitative relations*, which can be either (in the notation of Forbus ([6]):

  - *proportionalities*, where a parameter increases or decreases proportionally to another. Propotionalities correspond to equations.

  - *influences*, where a positive parameter causes another to increase or decrease. Influences correspond to differential equations.

The qualitative model can be represented as a graph. Figure 2 shows an example of a qualitative model.

Qualitative *reasoning* techniques can use such models to obtain qualitative predictions. These are based on *consistency* of qualitative value combinations with the proportionalities in the qualitative model. For example, in the model shown in Figure 2, it is impossible to have the combination $F = +, a = -$, since F is positively proportional to a.

It is also possible to qualitatively predict behavior over time. In this case, each combination of qualitative values consistent with the proportionalities forms a qualitative *state*. Influences are now used to rule out inconsistent transitions between qualitative states. For example, in the model of Figure 2, it is impossible to have a transition from:

$$x = +, F = -, a = -, v = -$$

4

to the qualitative state:

$$x = +, F = -, a = -, v = +$$

since this would contradict the positive influence between a and v: a being negative, v can only decrease, and thus not change signs from negative to positive. It turns out that in this example, qualitative reasoning is sufficient to uniquely identify a single successor to each state and thus correctly predict the oscillation of this system.

While qualitative reasoning is quite good at modelling the capabilities of an expert, it is very rare that the limited precision provided by the sign values is sufficient for a practically interesting problem. Numerically precise answers can be obtained by attaching interval constraints to parameters, as for example in `Q3` [2]. Using recently improved techniques of interval constraint propagation ([10, 5]), it is possible to obtain results with sufficient precision to be of practical use. It is also possible to use qualitative models for generating numerical simulation models, such as `SIMGEN` [7]. For differential equations, phase space techniques ([13, 14]) can offer more accuracy. However, these techniques require significantly more than purely qualitative knowledge, and thus may not be realistic in many cases.

3   Memory-based Reasoning: lacking structure

In its most general form, memory-based reasoning predicts the values of a *class* parameter given the values of a set of *attributes* by retrieving the previous case judged to be most similar and returning its class. This implies a metric for comparing similarities of different cases and thus comparing different combinations of matching attributes. In order to avoid excessive complexity, it is necessary to compose the metric from metric for individual attribute matches. This composition presents a number of problems:

- dependencies between attribute values: if attribute a_1 is equivalent to attribute a_2, they should not be counted twice when both match.

- dependencies between the influence and attribute has on the class: if the class $c = a_i - a_j$, then only matching a_i or a_j gives little information about the class. Only when both attributes match do we have an
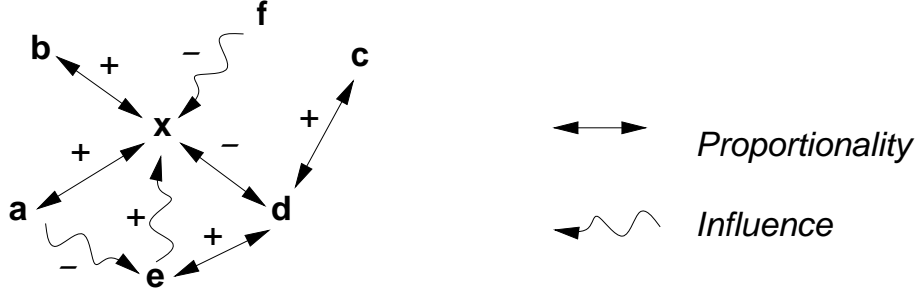
5

Figure 3: *Example of a qualitative model.*

indication that the class is also the same. Furthermore, there are many cases where the difference between attributes is identical but which would not match according to this criterion.

These are serious problems which are not easily solved. Many early memory-based reasoning systems simply used a weighted nearest-neighbour metric, where weights were chosen according to the statistical significance of the attributes. In [3], the authors report on experiments with a variety of different similarity metrics. These experiments showed clearly that other metrics were more powerful than the weighted nearest-neighbours approach. Other authors, for example Richter ([12]), attempt a more systematic treatment of similarity metrics based on theories of uncertain reasoning. While most of the approaches provide an approximate solution to the problem of dependencies between attribute values, they do little to address the dependencies which might exist in the way attributes influence the class.

4   The perfect couple

The interesting aspect of qualitative reasoning is that it provides tools for constructing accurate qualitative models of even rather complex systems. Even for devices which are quite poorly understood, experts are able to construct a fairly accurate qualitative model. This model can provide the structure which is lacking for accurate similarity measures in memory-based reasoning. Qualitative models offer a way to address both the dependencies between attribute values themselves as well as between attribute values and

the class. Assume that the qualitative model is represented as a graph as shown in Figure 3, and that the goal is to predict the value of parameter x given measurements of all others as well as a library of previous cases which give combinations of all attributes.

4.1 Instantaneous prediction

Consider first the case where prediction is instantaneous, i.e. the goal is to predict the value of x at the same time instant as the measurement of the attribute values. In this case, influence relations are to be ignored.

Dependencies of attribute values For evaluating the similarity of a case, only matches in parameters which have a direct influence on the parameter to be predicted are of importance. Thus, in the example of Figure 3, only matches in parameters a, b and d are significant. The degree of match in parameters c and e gives no further information about the similarity concerning x and thus should be ignored entirely. Thus, given the qualitative model, only parameters which are *directly* related to the class parameter are to be taken into account.

Dependent influence on the class When more than a single parameter is directly related to the class, their influence is dependent. Matching all parameters would ensure almost prefect prediction of the class, while matching a single one individually might give no information at all. In fact, unless the parameters have different orders of magnitude, it would usually be necessary to match all of them quite closely to ensure a good match of the class. It is unlikely that such close precedents can be found. However, using the signs assigned to qualitative proportionalities, it is possible to find cases which provide upper or lower *bounds* on the true value of the class x ([9]). In particular, assuming that p_c denotes the value of p in the case and p_0 the measured value of p, non-matching cases can be categorized into:

- *lower bound*: $(a_c \leq a_0) \wedge (b_c \leq b_0) \wedge (d_c \geq d_0)$

- *upper bound*: $(a_c \geq a_0) \wedge (b_c \geq b_0) \wedge (d_c \leq d_0)$

- irrelevant: all others

By ordering upper and lower bounds, we can construct a closest interval within which the value of x must lie, as long as the qualitative model is accurate and complete.

4.2   Prediction over time

Next, consider the case of prediction over time, i.e. to predict the value of the class at some time in the future given the current measurements of the class and related variables. Now, the influence of all parameters related to the class through proportionalities is already included in the degree to which the current value of the class matches. On the other hand, now the influences become important, since they describe how a parameter will change over time.

When the time interval over which prediction is made is sufficiently short, parameter values can be assumed approximately constant. The set of relevant parameters for the prediction can then be determined by a similar reasoning as above to be those which are directly influencing the class parameter. Thus, in the example of Figure 3, it is now the degree of match of parameters e and f together with the current value of x which determine the likelihood of the future value of x being the same.

Depending on the actual length of the time interval, changes in the parameters influencing the class may also have to be considered in the same way. For the example of Figure 3, this means that matching parameter a might also be important since it influences parameter e. Thus, the set of relevant parameters is now found as all those which are a certain maximal distance from the class in the graph of qualitative influences.

Once the set of relevant parameters has been determined, the same method as for instantaneous prediction can be applied to obtain bounds on the class value to be predicted. For parameters which are not directly influencing the class, the direction of the inequality is obtained by multiplying the influence signs on the path between it and the class.

4.3   Statistical correctness of predictions

If the qualitative model were completely accurate, i.e. it would include all relevant parameters, then the bounds obtained by the above methods would

be guaranteed. However, in practice any model leaves out certain parameters, which means that it cannot be trusted in limit cases. Furthermore, measurements may be inaccurate and falsify the result.

Hellerstein ([9]) already proposed a method whereby bounds are loosened to obtain a certain statistical confidence level in their validity. His method is based on not taking the closest bounding cases, but only the n-th closest, where n depends on the level of confidence to be achieved. In our applications, this has not improved the results, so I refer the reader to [9] for a detailed description of this method.

5 Experience with two practical examples

In my laboratory, we have applied the approach described above to two different applications:

- predicting several crucial parameters in the operation of a continous coffee roaster.

- predicting the final coffeine content in a decaffination process

I now describe the major results of each application in sequence.

5.1 Modelling a coffee roaster

Coffee roasting is an inherently unstable process and must be carefully controlled to avoid fires which destroy the entire load. The control problem was especially difficult for a new type of "continuous-roast" coffee roasters installed by Nestle in England and Japan. The numerical models which should form the basis for controlling the devices turned out the be highly inaccurate, resulting in frequent fires or shutdowns during the roasting process.

A process monitoring logs records snapshots of the system at thirty-second intervals. Each snapshot records the values of about 70 roaster parameters. Historical cases are provided in the form of one or more such process logs, each containing about 2600 snapshots (24 hours of recording). In order to give sufficient time for corrective action, the prediction should be made 10 minutes (20 samples) ahead. Thus, the monitoring logs were transformed into a case library where each case associates a set of observations with the value to be predicted 20 samples later.
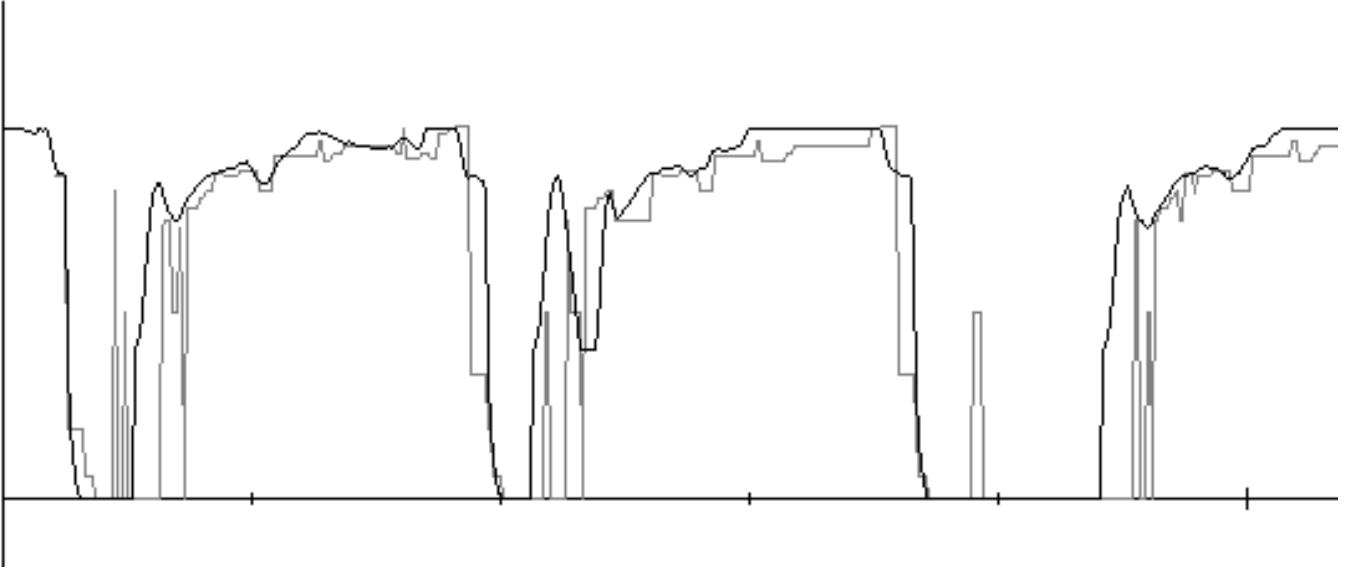
Figure 4: *A prediction obtained using the nearest-neighbor approach. The solid line shows the actual measurements, the dotted line is the prediction. Tics on the X-axis represent fifty-minute intervals.*

The qualitative model was based on structural schemata of the device as well as commonsense physical knowledge. It might seem desirable to involve an expert when formulating the model. However, it often turns out that an expert's understanding of a complex process like coffee roasting includes many misconceptions about how the process really works. The strength of an expert is *quantitative* experience with the process, which in this case is provided by the cases. Therefore, we consciously made not attempt to consult experts when constructing the model; it is incomplete and possibly contains some inaccuracies as well. By itself, it is insufficient for practically interesting predictions about the roaster.

Based on the indices as provided by the qualitative model, a simple nearest-neighbour approach can already provide rather accurate predictions. Figure 4 shows a sample prediction of the roaster's heater setting. Places where the parameter drops to zero indicate shutdowns, and we are looking at a particularly troublesome period. Note that the prediction is particularly bad when the roaster is restarted. This is easily explained since restart is
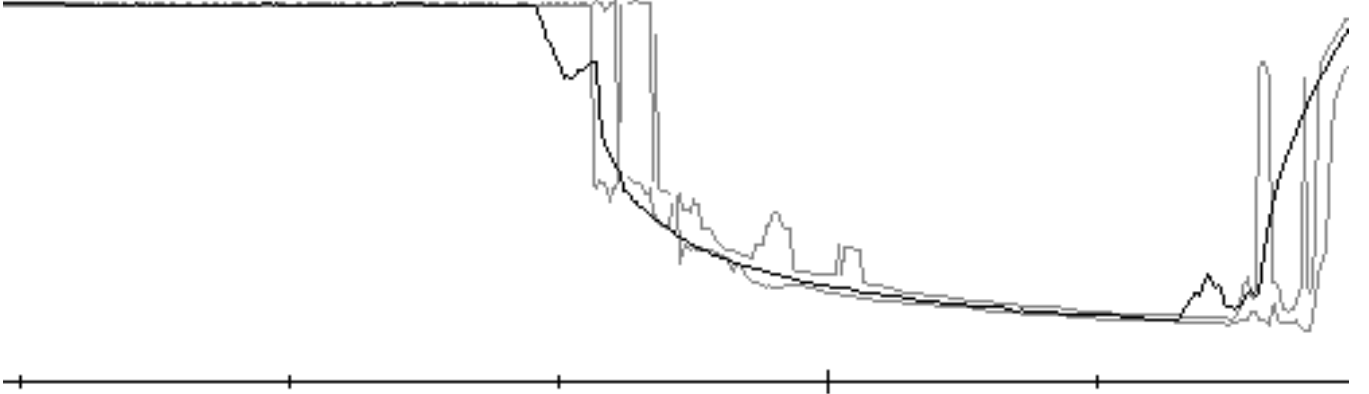
Figure 5: *Predicting bounds on the roaster temperature. The solid line shows the actual measurement, the dotted lines give the predicted bounds.*

done by operator action which is not represented in the cases!

However, the nearest-neighbour prediction fails to distinguish places where the confidence in the prediction is high from those where this confidence is low. Computing bounds on the actual value by classifying cases as upper and lower bounds provides such information. Figure 5 shows a sample prediction made using this approach, this time for predicting roaster temperature.

We selected five data files to use in all tests; logs from one roaster taken from five days in January 1994. In all cases, the roaster was producing the same recipe, which means that the external settings were identical. All logs contain one or more anomalous events; normally emergency shut-downs followed by some period of inactivity and a process restart. Depending on the test, we use between one and four logs as the source of historical cases (training examples), and one as a test set. Each data point in the graphs that follow represents the average of five test runs, one using each of the logs as a test set.

To verify our hypothesis that our approach degrades gracefully with the
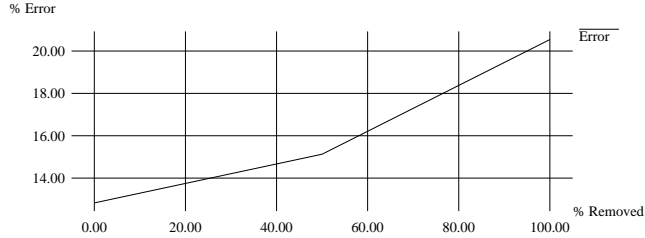
Figure 6: *Bounds widen slowly as the qualitative model degrades.*

quality of the input model, we ran a test in which we gradually removed constraints from the model (see Figure 6). This test was done using the bounding approach. Since the presence of large amounts of data tends to improve performance the performance of poor indexing schemes (given enough examples, even a poor indexing scheme will find an applicable example), we emphasized the model's importance by only using a single historical data set in each test run.

As constraints were removed from the model, the width of the bounds (graphed as "error") does indeed increase, but it does so gradually. Ultimately, when all constraints are removed, the only remaining index is the variable itself; prediction of the query parameter's future value is based solely on its past value, resulting in a confidence interval with an average width of more than 40% of the parameter's range (each bound is more than 20% distant from the center point).

The prediction using memory-based reasoning was the first to predict roaster behavior sufficiently accurate to be useful for human operators. Earlier attempts using neural networks had failed to achieve the required degree of accuracy. We should however note that Marc Goodman ([8]) has achieved results of similar quality on the same data; his system uses memory-based reasoning with an indexing scheme derived from statistics on the data. The main drawback of this approach is that it is likely to suffer from *overfitting*: attributing unduly high importance to parameters whose relation to the result is purely coincidental to the particular data set. Qualitative models avoid this phenomenon by the fact that they are based on scientifically established physical principles.

12

5.2 Modelling a decaffination process

The second example is a model of a decaffination process. This process, typical of many processes in the food industry, extracts the main elements of the solid coffee beans into a liquid, filters the liquid to extract the caffeine, and then reinserts it into the solid matter. It is important to predict as early as possible in the process whether the final caffeine contents will be sufficiently low to satisfy the quality requirements so that corrective action can be taken if necessary.

In order to make this prediction, we had available about 5'000[1] cases, each of which gives all parameters in the decaffination of a particular load of coffee beans as well as the final caffeine content. The prediction here is not a prediction over time, but of the caffeine content of the current load of coffee beans.

The qualitative model related all parameters of the process and allowed to restrict the attention to a small number which directly influence the caffeine content. Some of these parameters which proved statistically insignificant were not considered. Finally, only 6 parameters were selected as important for prediction. They are:

1. amount of total solid matter fed into the extraction cycle: positively related to caffeine contents.

2. activity index of filter in the extraction cycle: negatively related to caffeine contents.

3. total amount of filter material

4. amount of caffeine measured in the extraction cycle: positively related to caffeine contents.

5. extraction time

6. extraction temperature

It is interesting to note that these were not the parameters which were most significant in the overall statistics for predicting caffeine content. Even an

---

[1]The exact number varies between experiments because some do not contain all parameters.

expert in the plant operation required some time before being convinced that these parameters were in fact useful. Intuitively, this is because they are important only in certain combinations: suppose that the caffeine content c was the sum of two independent parameters a and b: $c = a + b$. If both a and b vary over a considerable range, c might end up being statistically entirely independent of a and b by themselves, in spite of being very much related to their sum. Note, however, that the most important parameter, the initial caffeine content of the beans before the process, is unavailable.

We again predicted both individual values using the nearest-neighbour method and intervals using the classification of cases into bounds. However, it turned out that when using all 6 parameters, the interval prediction very often failed to find enough cases to obtain significant lower or upper bounds. We thus reduced the set of parameters to (1), (2) and (4) only.

Figure 7 gives an idea of the prediction error using nearest-neighbour indexing based on the qualitative model. It shows the number of cases where the prediction error exceeded a certain bound. The cutoff value for accepting a batch of decaffinated coffee is about 0.05; while there are almost no cases which have that much error, there are quite many where the prediction error is of the same order of magnitude so that some improvements would still be desirable.

Figure 8 shows the width of the bounds computed using the qualitative relations. This gives an idea of how much precision can be expected from the method. Note in particular that the interval width was never greater than 0.05, whereas the maximum error in the nearest-neighbour method was about 0.07.

The overall prediction accuracy of both the nearest-neighbour and the interval-based approach is slightly better than what had been obtained with statistical tools and a neural network, but not good enough yet to be the basis for actual plant operation. It is likely that collecting more data will improve the situation by making it possible to use all important indexing parameters.

6   Conclusions

The most important problem in computing today is the software engineering bottleneck, which is to a large extent a knowledge engineering bottleneck.
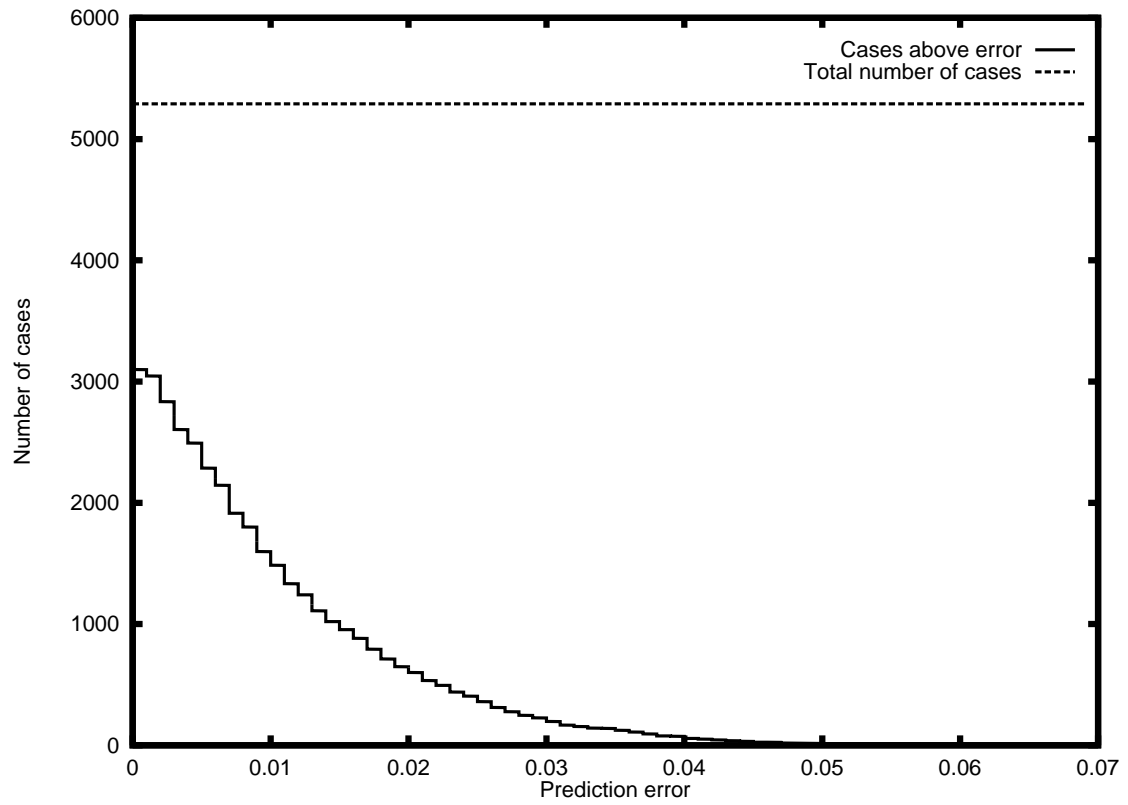
Figure 7: *Number of cases where the prediction error of final caffeine content exceeds the given limit, using nearest neighbour indexing with indices given by the qualitative model.*
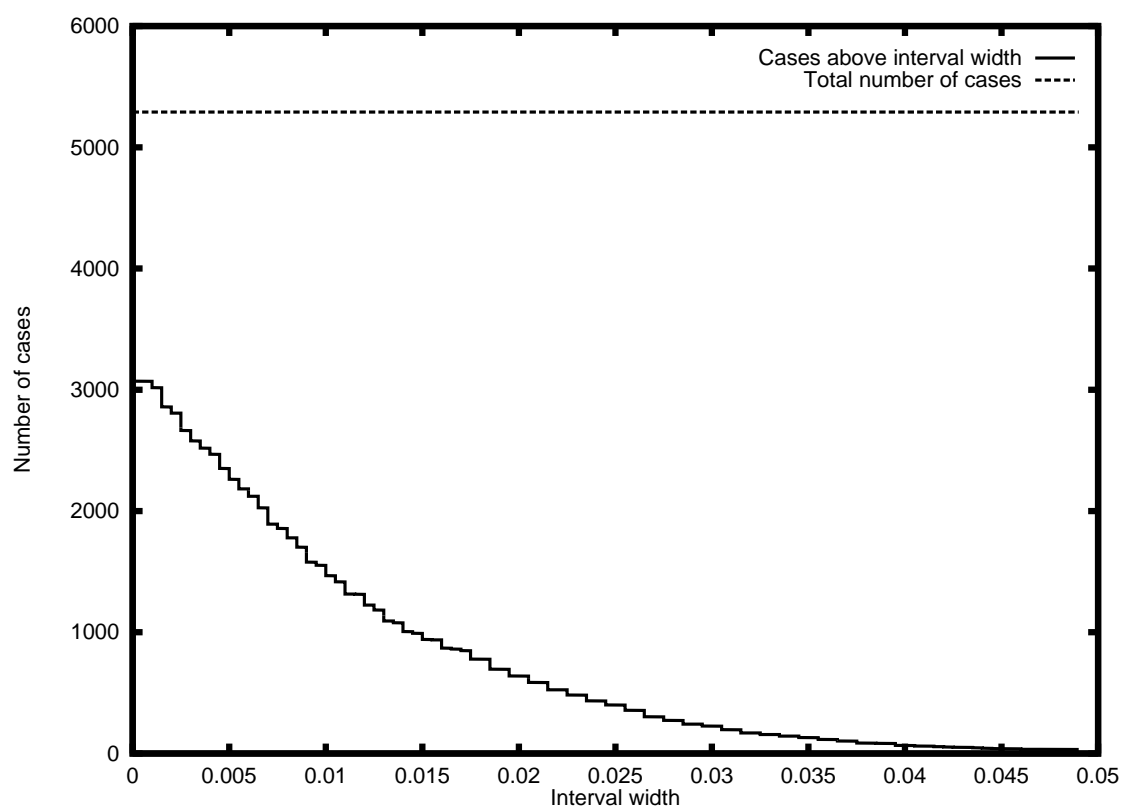
Figure 8: *Number of cases where the size of the prediction interval exceeds a certain threshold, using indexing by the qualitative relations.*

Ever decreasing hardware costs are making techniques for efficient knowledge acquisition all the more important.

While it would be most desirable to have computers which can learn the required knowledge fully autonomously, the complexity of practically interesting problems and the relative sparsity of data makes this goal difficult to reach. Automated learning, at least for now, requires some bootstrapping with human knowledge.

I think that the approach outlined in this paper is an interesting direction for resolving the knowledge engineering bottleneck. Both qualitative models and libraries of cases can be obtained with relatively little knowledge engineering effort. In combination, the two techniques neatly cover each other's shortcomings:

- the imprecision and ambiguities inherent in the qualitative model are irrelevant since the cases provide numerically precise answers. The qualitative model does not even have to be complete.

- the qualitative model provides sound and justified criteria for selecting cases, allowing subtleties which statistical criteria are likely to miss.

More advanced techniques of qualitative modelling, for example simulation ([11]) and envisionments ([6]), could extend these techniques to devices whose qualitative relationships vary over time. Also, similar techniques should be applicable in tasks other than prediction, such as diagnosis and design.

Acknowledgements

References

[1] **D. Aha, D. Kibler, and M. Albert**: "Instance-based learning algorithms," *Machine Learning* **6**, 1991.

[2] **D. Berleant, B. Kuipers**: "Combined qualitative and numerical simulations with Q3," in B. Faltings, P. Struss (eds.): *Recent Advances in Qualitative Physics*, MIT Press, 1992

[3] **R.H. Creecy, B.M. Masand, S.J. Smith, D.L.Waltz**: "Trading MIPS and Memory for Knowledge Engineering," *Communications of the ACM* **35**(8), August 1992

[4] **Johan de Kleer, John Seely Brown**: "A qualitative physics based on confluences," *Artificial Intelligence* **24**, 1984, pp. 7-84

[5] **B. Faltings**: "Arc-consistency for continous variables," *Artificial Intelligence* **65**, 1994, pp. 363-376

[6] **K. Forbus**: "Qualitative Process Theory," *Artificial Intelligence* **24**, 1984

[7] **K. Forbus, B. Falkenhainer**: "Self-Explanatory Simulations: Scaling up to large models," *Proceedings of the 10th National Conference of the AAAI*, AAAI/MIT Press, 1992

[8] **M. Goodman**: "Results on Controlling Action with Projective Visualization," *Proceedings of the 12th National Conference of the AAAI*, AAAI/MIT Press, 1994

[9] **J. Hellerstein**: "Obtaining Quantitative Estimates from Monotone Relationships," in B. Faltings, P. Struss (eds.): *Recent Advances in Qualitative Physics*, MIT Press, 1992

[10] **E. Hyvönen**: "Constraint reasoning based on interval arithmetic: the tolerance propagation approach," *Artificial Intelligence* **58**, 1992, pp. 71-112

[11] **Ben Kuipers**: "Qualitative Simulation," *Artificial Intelligence* **29**, 1986, pp. 289-338

[12] **M.M. Richter**: "On the Notion of Similarity in Case-Based Reasoning," in G. della Riccia et al (eds.): *Mathematical and Statistical Methods in Artificial Intelligence*, Springer Verlag 1995, pp. 171-184

[13] **E. Sacks**: "Automatic analysis of one-parameter planar ordinary differential equations by intelligent numerical simulation," *Artificial Intelligence* **48**, 1991, pp. 27-56

[14] **F. Zhao**: "Extracting and representing qualitative behaviors of complex systems in phase spaces," *Artificial Intelligence* **69**, 1994, pp. 51-92