Qualitative Modelling

Ivan Bratko

Faculty of Computer and Information Sc.,

University of Ljubljana

Abstract. Traditional, quantitative simulation based on quantitative models aims at producing precise numerical results as answers to user's questions about the problem domain. Such precise numerical answers are often overly elaborate, and they contain much more information than it is actually needed. In every day life, humans use common sense to reason about problems *qualitatively*, without numbers. In the area of Artificial Intelligence, methods exist for qualitative modelling and simulation. In this paper, we review some ideas of qualitative reasoning and modelling. We also discuss qualitative data mining as an approach to the analysis of numerical data, and Q2 learning which combines qualitative and quantitative approaches to modelling from data.

Keywords:
modelling, qualitative, quantitative, simulation, data mining, Q2 learning

Quantitative vs. qualitative modelling

Traditional, quantitative modelling and simulation give precise numerical answers. For everyday use, such answers are often overly elaborate. For example, consider the bath tub in Figure 1. Assume the tub is initially empty, there is a constant flow from the tap, and that the drain is closed, so there is no out flow. What will happen?

To answer this question, the physicist's solution would be to write down a differential equation model of this system, and run this model by numerical simulation. The numerical simulation would produce a table with, say, 1000 rows, giving the exact values of the level of water at consecutive tabulated time points. The table would show, for example, that the level will reach the top of the tub at 65.5 cm in 259.3 sec. For everyday use, such an elaborate answer is overkill. A common sense answer that completely suffices for everyday purposes, and is actually much more appropriate, is instead something like this: "The water level will keep increasing and will eventually reach the top. After this, water will be overflowing and cause a flood in the bathroom." This gives just a useful summary of a possibly large amount of quantitative information. The physicist's answer was *quantitative*, giving precise numerical information. The common sense answer was *qualitative*, just giving a useful summary of the large amount of quantitative information.

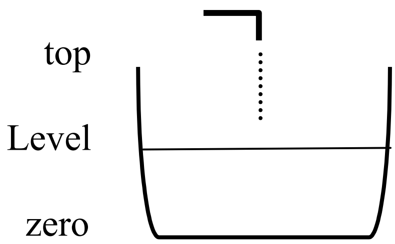

Figure 1: Bath tub with some input flow and closed drain.

Qualitative modelling and reasoning in AI

The humans are good at common sense, qualitative reasoning. Traditionally, computer-based methods are on the other hand mostly numerical -- just the opposite to common sense. Can the common sense, qualitative approach also be computerised? The area of qualitative reasoning and modelling in Artificial Intelligence aims at this (e.g. Weld and de Kleer 1990). It is concerned with the formalisation of and algorithms for qualitative reasoning about the world, producing qualitative, non-numerical answers to questions that are typically answered numerically by "proper" physics. To emphasise the contrast between the "proper" physics as taught in schools, and the qualitative, common sense reasoning about the physical world, the qualitative physics is sometimes also called *naive physics*.

Principles of qualitative abstraction

Qualitative reasoning is usually viewed as an abstraction of quantitative reasoning. Accordingly, in qualitative reasoning some numerical details are discarded; instead, a rather simpler qualitative summary of these numerical details is retained. There are many ways of abstracting away from detailed numerical information. Some abstraction principles are explained through examples in the following paragraphs.

Numbers are abstracted into symbolic values and intervals. For example, consider the quantitative statement: *Level* at 3.2 sec is 2.6 cm. A qualitative abstraction of this is: *Level* at time *t1* is between the bottom and the top of the bath tab. Notice here that 3.2 sec has been replaced by a symbolic time point *t1*. So instead of giving exact time, this just says that there is a time point, referred to as *t1*, at which *Level* has the given qualitative value. Regarding this qualitative value, the whole set of numbers between 0 and 65.5 has been collapsed into a symbolic interval *zero..top*. A further abstraction would be to disregard the top of the tub as an important value, and simply state: *Level* at time *t1* is positive, written as: $Level(t1) = pos$.

Another qualitative abstraction principle is to simplify the time derivatives into just *directions of change*. Such a qualitative statement can be: *Level* at time *t1* is increasing.

Another very useful qualitative abstraction principle is to simplify functions into *monotonic relations*. Consider for example the quantitative statement that states the relation between the amount of water and the level of water:

$$Amount = f(Level) = 25 * Level^2 + 45 * Level.$$

A qualitative abstraction of this can be: For $Level \geq 0$, *Amount* is a monotonically increasing function of *Level*, written formally as: $Amount = M^+(Level)$. That is: if *Level* increases then *Amount* increases as well, and vice versa. Notice that this statement is very simple, it is based on common everyday experience. The qualitative statement is also very general because it is true for *every* bathtub, regardless if its shape, it is actually true for every container.

Qualitative abstraction is related to *qualitative modelling*. Numerical models are an abstraction of the real world. Qualitative models are often viewed as a further abstraction of numerical models. In this abstraction some quantitative information is abstracted away. For example, a quantitative model of the water flow in a river may state that the flow *Flow* depends on the level *Level* of water in the river in some complicated way which also takes into account the shape of the river bed. In a qualitative model this may be abstracted into a monotonically increasing relation:

$M^+(\,Level, Flow)$

This just says that the greater the level the greater the flow, without specifying this in any more concrete and detailed way. Obviously, it is much easier to design such coarse qualitative models than precise quantitative models.

Why qualitative modelling?

Here we discuss some advantages of qualitative modelling with respect to the traditional, quantitative modelling. Of course there are many situations where a qualitative model, due to lack of precise numerical information, is not sufficient. However, there are many situations in which a qualitative model has advantages.

First, qualitative modelling is easier than quantitative modelling. Precise relations among the variables in the system to be modelled may be hard or impossible to determine, but it is usually still possible to state some qualitative relations among the variables. Also, even if a complete quantitative model is known, such a model still requires the knowledge of all the, possibly many, numerical parameters in the model. For example, a numerical physiological model may require the precise electrical conductance of a neuron, its length and width etc. These parameters may be hard or impossible to measure. Yet, to run such a numerical model, a numerical simulator will require the values of all these parameters to be specified by the user before the simulation can start. Usually the user will then make some guesses at these parameters and hope that they are not too far off their real values. But then the user will not know how far the simulation results are from the truth. The user will typically not know even if the obtained results are *qualitatively* correct. With a qualitative model, much of such guesswork can be avoided, and in the end the user will at least be sure about the qualitative correctness of the simulations. So, paradoxically, quantitative results, although more precise than qualitative results, are in greater danger of being incorrect and completely useless, because the accumulated error may become too gross. For example, in an ecological model, even without knowing the precise parameters of growth and mortality rates etc. for the species in the model, a qualitative model may answer the question whether certain species will eventually become extinct, or possibly different species will interchange their temporal domination in time cycles. A qualitative simulator may find such an answer by finding all the possible qualitative behaviours that correspond to all possible combinations of the values of the parameters in the model.

Another point is that for many tasks, numerical precision is not required. Often it only obscures the essential properties of the system. Generic tasks in which qualitative modelling is often more appropriate include functional reasoning, diagnosis and structural synthesis. *Functional reasoning* is concerned with questions like: How does a device or a system work? In a *diagnostic task* we are interested in defects that caused the observed abnormal behaviour of the system. Usually, we are only interested in those deviations from the normal state that caused a behaviour that is qualitatively different from normal.

The problem of *structural synthesis* is: Given some basic building blocks, find their combination which achieves a given function. For example, put the available components together to achieve the effect of cooling. In other words, invent the refrigerator from "first principles". The basic building blocks can be available technical components, or just the laws of physics, or materials with certain properties. In such design from first principles, the goal is to synthesise a structure capable of achieving some given function through some mechanism. In the early, most innovative stage of design, this mechanism is described qualitatively. Only at a later stage of design when the structure is already known, quantitative synthesis also becomes important.

The use of qualitative models requires qualitative reasoning. Some techniques of qualitative reasoning are presented e.g. in (Kuipers, 1993) and (Bratko 2001, chapter 20). Bratko et al. (1989) present a complex application of qualitative modelling in the diagnosis of cardiac arrhythmias.

Qualitative data mining

Consider the mining of numerical data. Usually, in machine learning used for numerical data mining, the goal is to construct a numerical function of the form $y = f(x_1, x_2, ...)$ where y is a distinguished variable called the dependent variable, and $x_1, x_2, ...$ are attributes (or independent variables). Examples of this kind of numerical learning are regression trees (CART, Breiman et al. 84), Retis (Karalič 92), M5 (Quinlan 1993).

In contrast to this, we may consider *qualitative* data mining which aims at finding *qualitative* patterns, or qualitative relationships in numerical data. An obvious motivation for qualitative data mining comes from the fact that for some tasks, qualitative models are more suitable than classical quantitative, numerical models, as discussed above. This kind of reasoning enables the solution of certain types of problems without resorting to numerical computation, and without the use of a quantitative model of the system in question. When a problem can be solved at the qualitative level of abstraction, there is no need for building a quantitative model – often a demanding or unrealistic task. Building qualitative models for complex systems should be easier.

While there are many machine learning or data mining tools that support the building of numerical models from data, there are few tools to support the building of qualitative models from data. Below we present one approach to qualitative data mining, realized in the program QUIN in which the target concepts are expressed by so-called *qualitative decision trees*.

The QUIN approach to qualitative data mining

QUIN (Qualitative Induction) is a learning program that looks for qualitative patterns in numerical data (Šuc 2001; Šuc and Bratko 2001). These patterns are then combined into a so-called *qualitative tree*. Induction of qualitative trees is similar to the well-known induction of decision trees. The difference is that in decision trees the leaves are labelled with class values, whereas in qualitative trees the leaves are labelled with what we call *monotonic qualitative constraints* (MQC). MQCs are a generalisation of monotonic relationships explained earlier, that are widely used in the field of qualitative reasoning, for example $Y = M^{+}(X)$. This says that Y is a monotonically increasing function of X. In general, MQCs can have more than one argument. For example, $Z = M^{+,-}(X,Y)$ says that Z monotonically increases in X and decreases in Y.

Monotonic constraints can be combined into if-then rules to express piece-wise monotonic functional relationships. For example:

if $X < 0$ then $Y = M^{-}(X)$ else $Y = M^{+}(X)$

Nested if-then expressions can be represented as trees, called *qualitative trees* (Šuc 2001). Qualitative trees are similar to regression trees (Breiman et al. 1984). Both regression and qualitative trees describe how a numerical variable depends on other variables. The difference between the two types of trees only occurs in the leaves. A leaf of a regression tree specifies a numerical regression function that tells how the class variable numerically depends on the attributes within the scope of the leaf. On the other hand, a leaf in a qualitative tree only specifies the relation between the class and the attributes *qualitatively*, in terms of monotonic qualitative constraints.

QUIN takes as input a set of numerical examples and looks for regions in the data space where monotonicity constraints hold. Such a set of qualitative patterns are represented in terms of a *qualitative tree*. As in decision trees, the internal nodes in a qualitative tree specify conditions that split the attribute space into subspaces. In a qualitative tree, however, each leaf specifies a MQC that holds among the input data that fall into that leaf. As a simple example consider a data set with three variables X, Y and Z where data triples (X, Y, Z) correspond to the function $Z = X^2 – Y^2$, possibly with some Gaussian noise added. When QUIN is asked to find in these data qualitative constraints on Z as a function of X and Y, QUIN generates the qualitative tree that can be represented by the following nested if-then-else expression:

```
if  X < 0  then
    if  Y < 0  then  Z = M-,+(X,Y)
               else  Z = M-,-(X,Y)
else
    if  Y < 0  then  Z = M+,+(X,Y)
               else  Z = M+,-(X,Y)
```

This tree partitions the data space into four regions that correspond to the four leaves of the tree. A different MQC applies in each of the leaves. The tree describes how Z qualitatively depends on X and Y. QUIN can tolerate some noise in the data and would induce in this example the same tree (with thresholds for X and Y slightly distorted) even when moderate noise is added to the data. Technical details of the QUIN algorithm can be found in (Šuc 2001) or (Šuc and Bratko 2001) where QUIN's performance on noisy data is also studied.

Q^2 learning

Q^2 learning (Šuc et al. 2004) stands for "qualitatively faithful quantitative learning". This approach combines qualitative and numerical learning from numerical data. The idea is to first find qualitative relationships in the data, and then use these qualitative properties to guide the numerical learning from the same data. This second stage can be carried out by a numerical regression method so that this method is forced to respect the observed qualitative relationships in the data. The Q^2 method as an approach to problem solving is illustrated in Fig. 2.

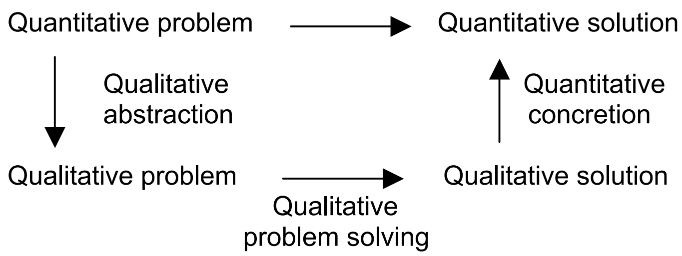

Fig. 2 Solving quantitative problems by means of qualitative abstraction

Advantages of Q^2 learning can be two fold. The induced qualitative model enables nice causal interpretation of the relations among the variables in the system, as one would expect from a qualitative model. More surprisingly, however, it was also shown in applications of Q^2 that the final quantitative model may yield numerical predictions that are significantly more accurate than those provided by state-of-the-art numerical modelling methods on their own.

In (Šuc et al. 2004) Q^2 was applied to induce a model of a complex, industrially relevant mechanical system (a car wheel suspension system). Thus the main message of this case study is that a combination of methods for qualitative and quantitative system identification methods has good chances to attain significant improvements over numerical system identification techniques, including techniques of numerical machine learning methods, such as regression trees, model trees, and locally weighted regression. Examples of application of Q^2 learning in ecological modelling are described by Vladušič et al. (2005; growth of plankton in Lake Glumsoe) and Žabkar et al. (2005; prediction of harmful ozone concentrations in Ljubljana and Nova Gorica).

Summary and conclusions

In this paper we reviewed some concepts in the qualitative approach to modelling, and some ideas of qualitative data mining or qualitative machine learning, and their combination with numerical machine learning exemplified by the Q^2 approach. We also reviewed some applications in which this approach was applied. These case studies demonstrate, as one would expect, that induced qualitative patterns are useful to facilitate the user's understanding of the domain of application, and to enhance the knowledge about the domain. However, less expected, we also found another very common application scenario of qualitative learning, when we are not just interested in qualitative relations, but in a concrete quantitative solution (e.g. making quantitative predictions, synthesising a numerical controller for a dynamic system, numerical system identification). This is facilitated by the Q2 method which combines qualitative and numerical learning.

Acknowledgement

The work described in this paper was in part done within the research programme Artificial Intelligence and Intelligent Systems supported by the Slovenian research agency ARRS, and in the European 5th framework project Clockwork, supported by the European Commission.

References

Bratko, I. (2001) *Prolog Programming for Artificial Intelligence*, 3rd edition. Pearson Education.

Bratko, I., Mozetič, I., Lavrač, N. (1989*) KARDIO: a Study in Deep and Qualitative Knowledge for Expert Systems*. Cambridge, MA: MIT Press.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. *Classification and Regression Trees.* Monterey, CA: Wadsford, 1984.

Karalič, A. Employing linear regression in regression tree leaves. *Proc. ECAI'92 (10th European Conference on Artificial Intelligence)*, pp. 440-441. Vienna 1992. Wiley & Sons.

de Kleer, J., Williams, B.C. (1991, eds.) *Artificial Intelligence Journal*, Vol. 51 (Special Issue on Qualitative Reasoning about Physical Systems II).

Kuipers, B.J. (1994) *Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge*. Cambridge, MA: MIT Press.

Šuc, D. (2001) *Machine Reconstruction of Human Control Strategies*. PhD Thesis, University of Ljubljana, Faculty of Computer and Information Sc., Ljubljana 2001. Also appeared as book in series Frontiers in Artificial Intelligence and Applications. IOS Press, Amsterdam, The Netherlands (2003).

Šuc, D., Bratko, I. (2001) Induction of qualitative trees. *Proc. Machine learning: ECML 2001*: (Lecture notes in artificial intelligence, Lecture notes in computer science, 2167) pp. 442-453. Berlin: Springer, 2001,

Šuc, D., Vladušič, D., Bratko, I., 2004. Qualitatively faithful quantitative prediction. *Artificial Intelligence Journal* **158**, 189–214.

Vladušič, D., Kompare, B., Bratko, I. (2005) Modelling phytoplankton in Lake Glumsoe with Q2 learning. Ecological Modelling (in press).

Weld, D.S., de Kleer, J. (1990*) Readings in Qualitative Reasoning about Physical Systems*, San Mateo, CA: Morgan Kaufmann.

Jure Žabkar, J., Rahela Žabkar, R., Vladušič, D. , Čemas D. , Šuc, D. and Bratko, I. (2005) Q^2 prediction of ozone concentrations. Ecological Modelling (in press).