Modelling of control skill by qualitative constraints

Dorian Šuc and Ivan Bratko
Faculty of Computer and Information Sciences
University of Ljubljana, Slovenia
{dorian.suc, ivan.bratko}@fri.uni-lj.si.

Abstract

Controlling a complex dynamic system, such as a plane or a crane, usually requires a skilled operator. Such a control skill is typically hard to reconstruct through introspection. Therefore an attractive approach to the reconstruction of control skill involves machine learning from operators' control traces, also known as behavioural cloning. In the most common approach to behavioural cloning, a controller is induced as a direct mapping from system states to actions. Unfortunately, such controllers usually lack typical elements of human control strategies, such as subgoals or desired trajectory, and do not replicate the robustness of the human control skill. In this paper we investigate a novel approach, whereby qualitative constraints are induced from an operator's control traces. These constraints define a qualitative control strategy - a qualitative model of the operator's skill. Such a qualitative control strategy defines a family of controllers and provides a space for controller optimization. Using the crane problem in a case study, this approach showed significant improvements over traditional approaches to skill reconstruction, both in terms of control performance and transparency of induced clones.

Introduction

Controllers for dynamic systems are traditionally designed by using methods of control theory, which assume the knowledge of the controlled system (a model). These methods work particularly well for linear systems, but have difficulties when the system's model is nonlinear, or is not known at all. In such cases, alternative approaches include the use of machine learning and/or rely on existing control skill of a human operator. This paper is a contribution in this direction.

Machine learning approaches to controller design, like reinforcement learning, genetic algorithms and neural networks typically don't use prior knowledge about the system to be controlled. This results in time-consuming experimentation with the dynamic system, low success rate of learning and lack of interpretability of learned controllers. Humans, however, rarely attempt to learn from scratch. They extract initial biases as well as strategies from their prior knowledge of the system or from demonstration of experienced operators.

The idea of *behavioural cloning* (a term introduced by Donald Michie (Michie 1993)), but probably first carried out by Donaldson (Donaldson 1964)) is to make use of the operator's skill in the development of an automatic controller. A skilled operator's control traces are used as examples for machine learning to reconstruct the underlying control strategy that the operator executes subconsciously. The goal of behavioural cloning is not only to induce a successful controller, but also to achieve better understanding of the human operators subconscious skill (Urbančič & Bratko 1994). The first such rule-based acquisition of real-time control skill was in (Chambers & Michie 1969). Behavioural cloning was later successfully used in problem domains as pole balancing, production line scheduling, piloting (Sammut *et al.* 1992) and operating cranes. These experiments are reviewed in (Bratko, Urbančič, & Sammut 1998). Controllers were usually induced in the form of decision or regression trees.

Although such clones do provide some insight into the control strategy, they in general lack conceptual structure that would clearly capture the causal relations in the domain and the goal structure of the control strategy. Ignoring causality and the dynamics of the system usually results in the following problems that have been observed with clones in the form of trees:

- Typically, induced clones are brittle with respect to small changes in the control task.
- The clone induction process typically has low yield: the proportion of successful controllers among all the induced clones is low, typically well below 50%.
- Resulting clones are purely reactive and inadequately structured as conceptualizations of the human skill. They lack typical elements of human control strategies such as goals, subgoals, phases and causality.

In this paper we use a new approach to behavioural cloning which handles significant nonlinearities and also enables qualitative treatment of human control strategy. The trajectory the operator is trying to follow is generalized separately from the system's dynamics and

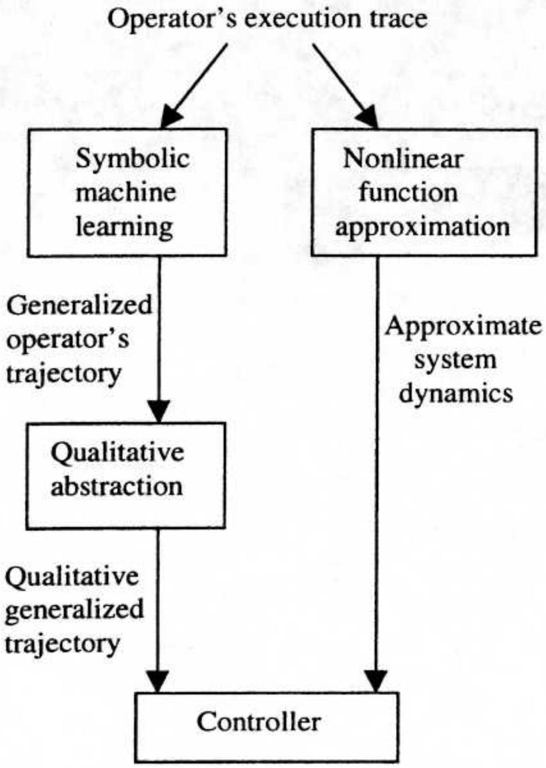

Figure 1: Reconstructing operator's control skill from his execution trace.

can be viewed as a *continuous subgoal*. In particular, we do not learn the trajectory in time, but rather the constraints among the state variables in the execution trace. These constraints determine the corresponding desired trajectory to the goal, also for system states other than those explicitly mentioned in the operator's execution trace. Actions that maintain the desired trajectory are computed using knowledge of the system's dynamics, learned by nonlinear function approximators. Our experiments performed in the crane domain (Šuc & Bratko 1999) and experiments in the Acrobot domain (Šuc & Bratko 1998) demonstrated that this approach significantly improves the yield of the cloning process and provides a good insight in the operator's subcognitive skill. In this paper we show that qualitative strategy, generalized from the operator's trajectory, is comprehensible and offers the possibility to optimise the operator's control strategy.

The structure of the paper is as follows. First we give a general description of our approach. The following two sections present the domain of container cranes and experiments where we used our approach to reconstruct the human crane control skill. In the next section we generalize the induced strategy into a comprehensible qualitative strategy. By transforming this qualitative strategy into operational quantitative strategies, we show that the qualitative strategy is general and successful. We also test robustness of the resulting controllers. Finally, we discuss some points of interest and give conclusions.

Our approach to skill modelling

General idea

The main idea of our approach (see Fig. 1) is to generalize the trajectory the operator is trying to follow and separately learn the system dynamics by a nonlinear function approximator. The learned nonlinear model of system dynamics is then used to compute control action which achieves the desired next state on the trajectory. When generalizing a given operator's trajectory, we do not learn the trajectory in time, but rather induce constraints between the state variables in the operator's execution trace. In this paper we investigate in particular the case where constraints are qualitative. Such constraints give rise to a *qualitative* control strategy.

These constraints can be used to control the system as follows. Ideally, to mimic the operator's control strategy, the constraints should hold in every state. So an appropriate control action should minimize the deviation of the predicted next state from the constraints. A measure of such deviation of a state from the constraints will be called *constraints error*. A constraints error simply measures the degree to which a system's state does not satisfy the constraints. Given a system state and the constraints, an appropriate action can be computed in many ways. In general it requires a model of the system's dynamics and minimization of the constraints error over the possible actions. One possibility, used also in this paper, is to induce constraints between the current and the next state in the execution trace and choose the action which minimizes error w.r.t. constraints in the next state.

One motivation for this approach to skill modelling is in the possibility of explaining the operator's control strategy. Induced symbolic constraints that characterize the operator's strategy can be used to explain such a subcognitive strategy. As shown later in the paper, induced constraints can be turned into *qualitative control rules*, which can give additional insight into the operator's control strategy and offers a possibility to optimise the operator's strategy.

Qualitative generalization of the operator's trajectory

The operator's trajectory is generalized by finding constraints among the state variables. In this paper constraints are induced from the operator's trace in two stages. First, constraints in the form of differential equations are induced. Then these equations are abstracted into qualitative constraints. This abstraction is explained in detail later in the paper. Here we explain briefly how differential equations were induced.

To induce constraints in the form of ordinary differential equations, we used the machine learning program called GoldHorn (Križman, Džeroski, & Kompare

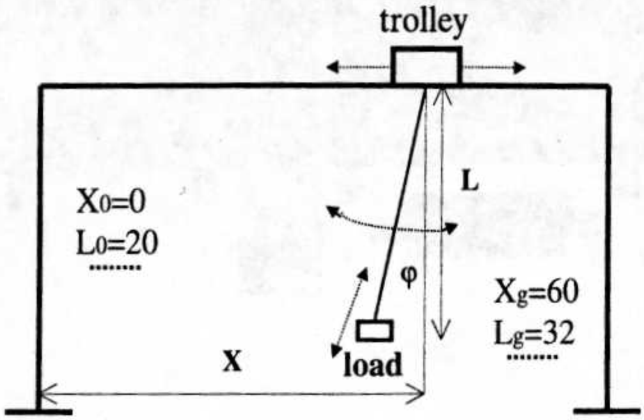

Figure 2: Container crane: the state of the system is specified by six variables: trolley position X and its velocity $\dot{X}$, rope inclination angle ϕ and its angular velocity $\dot{\phi}$, rope length L and its velocity $\dot{L}$. The system is controlled through force to the trolley in the horizontal direction and force in the direction of the rope. The task is to transport the load from its start position ($X_0 = \dot{X}_0 = \phi_0 = \dot{\phi}_0 = 0$, $L_0 = 20$, $\dot{L}_0 = 0$) to the goal position ($X_g{=}60, \dot{X}_g{=}\phi_g{=}\dot{\phi}_g{=}0$, $L_g{=}32$, $\dot{L}_g{=}0$).

1995). Given the behavior of the system, i.e. a sampled execution trace, GoldHorn attempts to find a set of differential equations that describe the dynamics of the system. GoldHorn does not simply fit the parameters of equations of given forms, but it also constructs new forms of equations. To do this, GoldHorn first introduces new terms by repeatedly applying operators, such as multiplication, to the state variables and their time derivatives. Then, given the set of all the terms, including the original variables and the newly constructed terms, differential equations are generated from these terms using linear regression.

The generalized operator's trajectory is induced by GoldHorn from one or more execution traces. GoldHorn induces constraints among the state variables in the form of differential equations and ranks them according to their significance (error estimates). One or more of the most significant induced equations are used as constraints which define the generalized operator's trajectory.

Learning the system dynamics

The system's dynamics can be learned by any nonlinear function approximator from some execution traces. We use locally weighted regression (Cleveland 1979; Schaal & Atkeson 1994), since it enables incremental learning and provides local linear model of the system's dynamics near the current state. Locally weighted regression is a kind of memory based learning, so no generalization is actually done in the training phase. We just need to store the observed points in the state space. In the prediction phase, points in the state space are weighted according to the distance from the query point and locally weighted linear model is computed. After answering the query the local model is discarded and a new locally weighted model is created to answer the next query. When the system is controlled, each current state is simply stored, and this experience can be used to predict the system's behavior in nearby points in the state space.

Container crane

To transport a container (see Fig. 2) from the shore to a target position on the ship, two operations are to be performed: positioning of the trolley, bringing it above the target load position (X_g), and rope operation, bringing the load to the desired height (L_g). The performance requirements include basic safety, stop-gap accuracy and as high capacity as possible. The last requirement means that the time for transportation is to be minimized. Consequently, the two operations are to be performed simultaneously. The most difficult aspect of the task is to control the swing of the rope. When the load is close to the goal position, the swing should ideally be zero.

A crane simulator was used in our experiments. The parameters of the system (lengths, heights, masses, etc.) are the same as those of the real cranes in Port of Koper. The state of the system is specified by six variables: trolley position X and its velocity $\dot{X}$, rope inclination angle ϕ and its angular velocity $\dot{\phi}$, rope length L and its velocity $\dot{L}$. Two control forces are applied to the system: force to the trolley in the horizontal direction and force in the direction of the rope.

We used experimental data from manually controlling the crane from a previous study (Urbančič & Bratko 1994). In that study, six students volunteered to learn to control the simulator. Remarkable individual differences were observed regarding the characteristics of the strategy they used. Some operators tended towards fast and less reliable operation, others were more conservative and slower, in order to avoid large rope oscillations. In experiments with behavioural cloning the main problem for regression trees was the swing control, i.e. large rope oscillations when the trolley approached its goal position. For this reason slower, more conservative traces were found by far most useful for cloning with regression trees. On the other hand, strategies of faster operators are more complex, requiring the exact timing and skill to decrease large rope oscillations due to large trolley accelerations. In this paper we use traces of the fastest operator, that is the operator who achieved shortest times. He was able to afford large initial swing caused by large acceleration, and later skillfully reduce the swing.

Skill reconstruction in the crane domain

Reconstruction of human control skill follows the general idea of our approach. Here we give details specific to the crane domain.

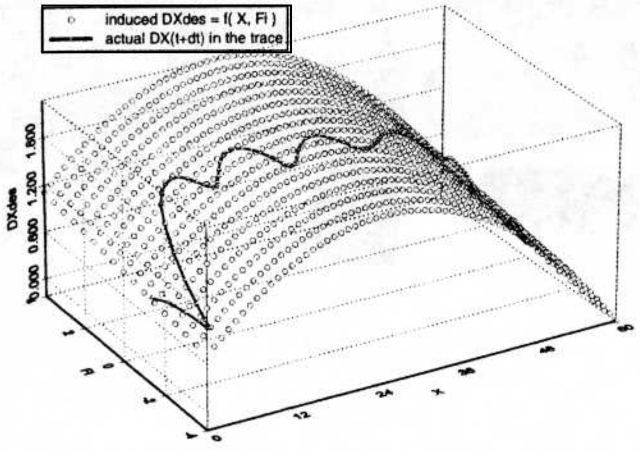

Figure 3: Induced quantitative strategy: the dotted surface represents the induced quantitative strategy (eq. 1) DX_{des} as a function of X and ϕ. The line represent the actual $DX(t+dt)$ from the original operator's trace. Qualitative rules (4) $Q_+(DX_{des}, \phi)$ and the rule *if* $X < 25.62$ *then* $Q_+(DX_{des}, X)$ *else* $Q_-(DX_{des}, X)$ can be observed.

First the generalized trajectory the operator is trying to follow is induced from one or more traces of the same subject. In the case of the crane control, two forces are applied to the system: the force to the trolley XF and the force to the rope YF. So the generalized trajectory consists of two variables: the desired trolley velocity DX_{des} and the desired rope length L_{des}. Of course, generalized trajectory could be expressed in a different way, for example (DX_{des},DL_{des}). We decided for (DX_{des},L_{des}) trajectory since it is easy to understand. The learned trajectory consists of DX_{des} and L_{des} as a function of other state variables, that is DX and L in the next state as the function of the current state.

When the trajectory is known, the action which achieves the desired next state on the trajectory can be computed using a locally linear model. So, the actual controller consists of two parts: nonlinear model of system dynamics and a desired trajectory. At each time step the controller considers the current state, estimates local linear model around the current state and computes action pair (XF,YF) which makes one step towards the desired next state on the trajectory.

This approach to skill reconstruction turned out to be generally very effective. The induced control strategy often concludes the task faster than the original operator's traces. Usually a more successful trajectory was induced from a faster execution trace. So individual differences between the operators, like differences in operator's speed and skill to decrease the swing of the rope, were reflected in the generalized trajectories induced from their traces. Other details of the experiments are given in (Šuc & Bratko 1999).

Induced quantitative strategy

Here we present a quantitative strategy learned from one of the traces of the fastest operator. The most significant equation, induced by GoldHorn from the operator's trace, was used as the desired trajectory of the trolley velocity (see Fig. 3):

$$DX_{des} = 0.902 - 0.0018X^2 + 0.090X + 0.050\phi \quad (1)$$

To control the rope, a very simple equation with higher error estimate was used:

$$L_{des} = 0.0037X^2 + 17.46$$

The equation in simple, but not accurate enough, since it states that L_{des} at the goal position X_g is 30.7. To conclude the task, the rope length at the goal position should be $L_{des}(X_g)$=L_g=32. This can be corrected by adding this difference to the learned L_{des}. In this way we get the rule to control the rope length:

$$L_{des} = 0.0037X^2 + 17.46 + 1.3 \quad (2)$$

The rules given by equations 1 and 2 describe the generalized operator's trajectory. The actual controller consisting of the generalized trajectory and the learned system dynamics was tested on the simulator and is faster than the original operator's trace. By inducing a successful controller only one goal of behavioural cloning was achieved. The motivation for introducing qualitative strategy is in better understanding of human control strategy and in the possibility of its improvement.

Qualitative strategy

Here we generalize the induced quantitative strategy into a qualitative strategy. We show that this qualitative strategy is easy to understand general (it is not sensitive to the exact settings of the qualitative parameters) and very successful. Since controlling the trolley is considerably harder than controlling the rope, we focus on DX_{des} trajectory. In all experiments we used L_{des} trajectory given by rule 2. Of course, the same approach as with DX_{des} trajectory could be used to generalize L_{des} trajectory into a qualitative strategy.

One way to express a qualitative strategy is by two qualitative proportionality predicates Q_+ and Q_- as defined by Forbus (Forbus 1984). The notation $Q_+(y,x)$ means that y is a function of a set of variables, including x, so that y is monotonic increasing (strictly increasing) in its dependence on x:

$$Q_+(y,x) \Leftrightarrow \frac{\partial y(x,...)}{\partial x} > 0 \quad (3)$$

We say that y is *positively related* to x. The representation $Q_+(y,x)$ allows to conjecture that the qualitative behavior of y cannot be controlled without taking into account the qualitative behavior of x and furthermore, positive (or negative) changes in x will be manifested in changes of y that are more positive (or negative) than they would have been without the change of x. The meaning of $Q_-(y,x)$ is analogous but in the opposite direction.

Note that this is a weaker relation than the other well known qualitative relation $M_+(y,x)$. $Q_+(y,x)$ only states that when x rises, y will also, barring other changes.

Qualitative constraints, defining a qualitative strategy, can now be derived by qualitative abstraction of the quantitative strategy given by rule 1:

$$\frac{\partial DX_{des}}{\partial \phi} = 0.050 > 0$$
$$\Rightarrow Q_+(DX_{des},\phi)$$
$$\frac{\partial DX_{des}}{\partial X} = -2X + 51.24$$
$$\Rightarrow Q_+(DX_{des},X) \text{ if } X < 25.62$$
$$Q_-(DX_{des},X) \text{ if } X > 25.62$$

These rules describe the qualitative control strategy:

$$\begin{array}{l} Q_+(DX_{des},\phi) \\ \text{if } X < X_{mid} \text{ then} \\ \quad Q_+(DX_{des},X) \\ \text{else } Q_-(DX_{des},X) \end{array} \tag{4}$$

This qualitative strategy provides good insight in the operator's control strategy. Actually, it gives all the qualitative knowledge, needed to bring the trolley above the goal position and complete the crane control task. The most interesting is the qualitative rule $Q_+(DX_{des},\phi)$. It is very important part of the operator's strategy and describes his skill to decrease the swing of the rope. It states the trolley's velocity should increase when the rope angle increases, and vice versa. By acceleration of the trolley when the rope angle increases and deceleration when the rope angle decreases, the angular velocity is decreased. This rule efficiently controls the swing of the rope, which is the crucial problem for all, except the most experienced operator (that is the operator, whose trace we used to learn the strategy). The second rule just says: increase the trolley's velocity (when the trolley is near its start position: $X < X_{mid}$) and decrease it later, to stop at the goal position. Note that X is increasing, as the trolley approaches the goal, so the rule $Q_+(DX_{des},X)$ ($Q_-(DX_{des},X)$) recommends to increase (decrease) the trolley's velocity.

The qualitative strategy was derived by qualitative abstraction of the induced quantitative strategy. The qualitative rules can also be observed in the induced quantitative strategy (see Fig. 3), which is a single special case of the more general quantitative strategy. This special case of the qualitative strategy successfully completes the task. In the rest of the paper we investigate if other strategies satisfying qualitative constraints are also successful, that is how general is the qualitative strategy, and can it be used to optimise the operator's control strategy.

One form of qualitative rule for $DX_{des}(X,\phi)$, consistent with 4, is:

$$\begin{array}{l} \text{if } X < X_{mid} \text{ then} \\ \quad DX_{des} = M_x^+(X) + M_{\phi 1}^+(\phi) \\ \text{else } DX_{des} = M_x^-(X) + M_{\phi 2}^+(\phi) \end{array} \tag{5}$$

M_x^+ (M_x^-), $M_{\phi 1}^+$ and $M_{\phi 2}^+$ are arbitrary strictly increasing (decreasing) functions for $X < X_{mid}$ ($X > X_{mid}$). The qualitative strategy, given by rule 5, defines a set of quantitative strategies S with free parameters X_{mid}, M_x^+, M_x^-,$M_{\phi 1}^+$ and $M_{\phi 2}^+$. To transform the qualitative strategy into an operational quantitative strategy, the qualitative parameters have to be concretized into quantitative values or functions. For example, the qualitative parameters of the quantitative strategy, given by rule 1, are:

$$X_{mid} = 25.62$$
$$M_{\phi 1}^+(\phi) = M_{\phi 2}^+(\phi) = 0.050\phi$$
$$M_x^+(X) = M_x^-(X) = 0.902 - 0.0018X^2 + 0.090X$$

Transforming qualitative strategy into operational quantitative strategies

The induced qualitative strategy, given by rule 5, defines a set of quantitative strategies S with free qualitative parameters X_{mid}, M_x^+, M_x^-, $M_{\phi 1}^+$ and $M_{\phi 2}^+$. Here we investigate how successful and general is the learned qualitative strategy, that is how sensitive it is to changes of the qualitative parameters. In order to do this we performed two sets of experiments:

- In the first experiment we used just basic knowledge of the control task and randomly generated functions satisfying the qualitative constraints. Results show that the qualitative strategy is general and that the exact selection of the qualitative parameters is not crucial for its success.
- In the second experiment we restricted qualitative functions to simple functions, and used knowledge of the control task to find constraints among parameters of the quantitative strategy. Experimentally it is confirmed that, by using some knowledge of the control task, it is easy to find the constraints among parameters of the quantitative strategy which filter out bad strategies. In this way, very successful strategies can be obtained. Since a generalized trajectory is expressed in a symbolic way, it is easy to understand, analyze and correct.

Randomly generated qualitative functions

Here we use basic knowledge of the control task and randomly generated functions with the required qualitative properties, to transform the induced qualitative strategy into operational quantitative strategies. Knowledge used consists of the task limits ($|\phi| < \phi_{max}$, $|DX_{des}| < DX_{max}$) and two simple facts of the control task:

1. the trolley starts towards the goal: $DX_{des}(0,0) \in (0, DX_{max})$
2. the trolley stops at X_g: $DX_{des}(X_g, 0)=0$

First we define a set of strictly increasing (decreasing) functions and use them to incorporate the knowledge of the control task into a set of quantitative strategies, based on induced qualitative strategy. Then experimental results are presented.

Let us define a set of functions F^+ (F^-), such that any $f^+ \in F^+$ ($f^- \in F^-$) is strictly increasing (decreasing) function defined on $[0,1]$ and mapping to $[0,1]$:

$$f^+, f^- : [0,1] \to [0,1]$$
$$f^+(x_1) > f^+(x_2) \Leftrightarrow x_1 > x_2$$
$$f^-(x_1) < f^-(x_2) \Leftrightarrow x_1 > x_2$$

Similarly we define a set of functions H^+ (H^-), such that any $h^+ \in H^+$ ($h^- \in H^-$) is strictly increasing (decreasing) function defined on $[-1,1]$ and mapping to $[-1,1]$, with the property that $h^+(0) = h^-(0) = 0$.

Different functions f^+ can be generated by taking $R+1$ random numbers sorted in increasing order $0 \leq y_i \leq 1$, $i = 0, 1, \ldots, R$ ($y_i < y_{i+1}$) and set $f^+(i/R)$ to y_i. The values at other points are linearly interpolated: $f^+(x) = y_j + (y_{j+1} - y_j)(Rx - j)$ where $j = \lfloor xR \rfloor$. Functions h^+ can be generated similarly.

The basic knowledge of the control task can be incorporated into a set of quantitative strategies based on qualitative strategy (rule 5) as follows. Since $|\phi| < \phi_{max}$, increasing functions $M^+_{\phi 1}$ and $M^+_{\phi 2}$ can be written as:

$$M^+_{\phi 1, \phi 2}(\phi) = k_1 h^+_{1,2}(\frac{\phi}{\phi_{max}})$$

Since $|DX_{des}| < DX_{max}$, $k_1 \in (0, DX_{max})$. Because the trolley must start towards goal ($DX_{des}(0,0) > 0$) and $|DX_{des}| < DX_{max}$ and $M^+_x(X)$ is increasing function, M^+_x can be written as:

$$M^+_x(X) = DX_{max} f^+_1(X/X_{mid})$$

Since the trolley must stop at the goal position ($DX_{des}(X_g, 0)=0$) $M^-_x(X)$ can be written as:

$$M^-_x(X) = DX_{max} h^-_3(\frac{X - X_g}{X_g - X_{mid}})$$

This gives a set of quantitative strategies $S_{q0} \subset S$:

if $X < X_{mid}$ then

$$DX_{des} = DX_{max} f^+_1(\frac{X}{X_{mid}}) + k_1 h^+_1(\frac{\phi}{\phi_{max}})$$

else

$$DX_{des} = DX_{max} h^-_3(\frac{X - X_g}{X_g - X_{mid}}) + k_1 h^+_2(\frac{\phi}{\phi_{max}})$$

$X_{mid} \in (0, X_g)$, $k_1 \in (0, DX_{max})$ (6)

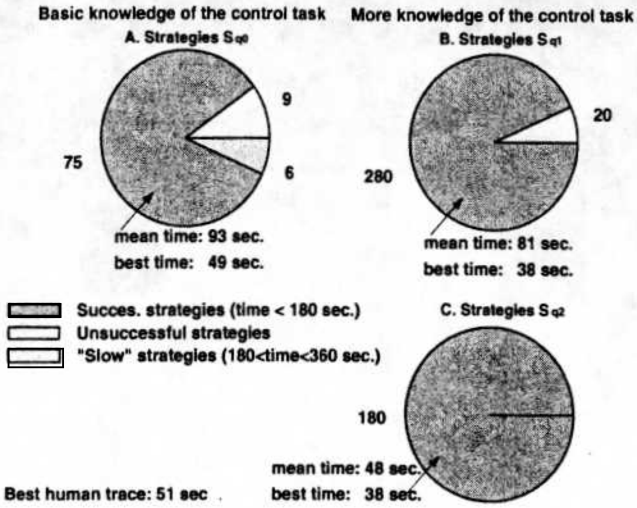

Figure 5: Efficiency of the qualitative strategy: Pie A. gives results with randomly generated qualitative functions. Out of 90 different quantitative strategies 75 finish the task in 180 sec. or less. Another 6 are successful in 360 sec. Pies B. and C. give results with qualitative functions using knowledge of the control task. All of 180 strategies from S_{q2} are successful and very fast. Even their mean time (48 sec.) is better than the time of the fastest human trace (51 sec.). Unsuccessful strategies from S_{q0} and S_{q1} approach the goal position, but are too slow to complete the task in 180 seconds.

These quantitative strategies were tested on the simulator. X_{mid} was set to values 15, 30, 45 and k_1 to values $DX_{max}/2$, $DX_{max}/4$ and $DX_{max}/8$. For each of these parameter settings, we generated 10 sets of random functions $f^+_1, h^+_1, h^+_2, h^-_3$ with $R = 10$. In this way we get a set of $3 \times 3 \times 10 = 90$ different quantitative strategies referred to as S_{q0}. Considering the fact that we used a qualitative strategy, where qualitative functions were random and we used only the basic knowledge of the crane task, results are very good. Out of 90 different quantitative strategies 75 finished the task in 180 sec. or less, some of them faster than 50 seconds. The details are given on Fig. 5.

Using knowledge of the control task

Here we restrict qualitative functions to linear and quadratic functions and use the knowledge of the control task to find constraints among parameters of the quantitative strategy. In order to do this, we define quantitative strategies S_{q1} as a subset of strategies S with the properties:

1. *quadratic M^+_x and M^-_x and linear $M^+_{\phi 1}$ and $M^+_{\phi 2}$.* Qualitative parameters, i.e. functions $M^+_{\phi 1}(\phi)$ and $M^+_{\phi 2}(\phi)$ are restricted to linear functions $k\phi + n$ and functions $M^+_x(X)$ and $M^-_x(X)$ are restricted to quadratic (or linear, when p=0) functions $px^2 + rx + m$. Parameters n and m are (without loss of gener-

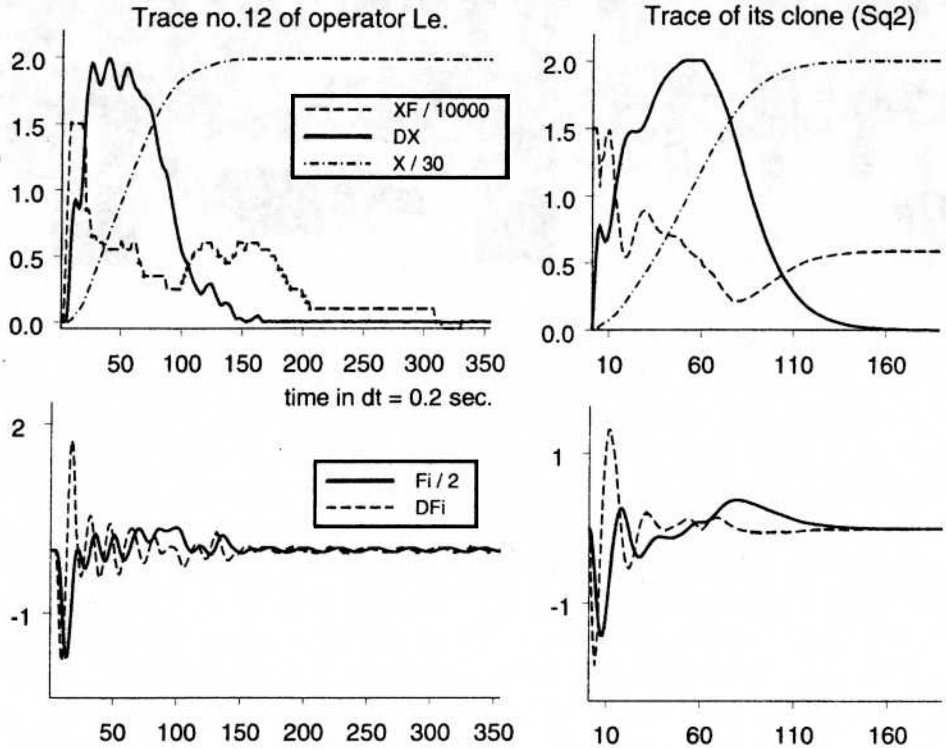

Figure 4: Original and its clone: on the left is the original trace and on the right is the trace of its clone, using one of the controllers resulting from the learned qualitative strategy. The original concludes the task in 72 sec. The clone is much faster and concludes the task in 38 sec. Decreasing of the rope swing can be observed in both traces.

ality) joined into DX_{mid}. This defines a subset of quantitative strategies of the form:

$$\begin{aligned} &\text{if } X < X_{mid} \text{ then} \\ &\quad DX_{des} = DX_{mid1} + p_1 x^2 + r_1 x + k_1 \phi \\ &\text{else } DX_{des} = DX_{mid2} + p_2 x^2 + r_2 x + k_2 \phi \end{aligned} \tag{7}$$

2. *Satisfy the task requirements* (obey task limits, stay at the goal position).

3. *Start towards goal* ($DX_{des}(0,0) > 0$).

4. *Smooth and fast at* X_{mid} ($DX_{des}(X_{mid} - \epsilon, 0) = DX_{des}(X_{mid} + \epsilon, 0) = DX_{max}$).

Each of these properties imposes constraints on the parameters. For example, the property that the strategy obeys task limits ($\phi < |\phi_{max}|, DX < |DX_{max}|$) requires $-DX_{max} < k_1\phi, k_2\phi < DX_{max}$, yielding $-\frac{DX_{max}}{\phi_{max}} < k_1, k_2 < \frac{DX_{max}}{\phi_{max}}$. In this way we derive constraints on the parameters of the quantitative strategies S_{q1} (rule 7):

$$\begin{aligned} &0 < X_{mid} < X_g \;,\; 0 < DX_{mid1} < DX_{max} \\ &-\frac{DX_{max}}{\phi_{max}} < k_1 = k_2 < \frac{DX_{max}}{\phi_{max}} \\ &-\frac{DX_{max} - DX_{mid1}}{X_{mid}^2} < p_1 < \frac{DX_{max} - DX_{mid1}}{X_{mid}^2} \\ &-\frac{DX_{max}}{(X_g - X_{mid})^2} < p_2 < \frac{DX_{max}}{(X_g - X_{mid})^2} \\ &r_1 = \frac{DX_{max} - DX_{mid1}}{X_{mid}} - p_1 X_{mid} \\ &r_2 = \frac{DX_{max}}{X_{mid} - X_g} - p_2(X_{mid} + X_g) \\ &DX_{mid2} = -X_g(X_g p_2 + r_2) \end{aligned} \tag{8}$$

Further we define strategies S_{q2} as a subset of strategies S_{q1} with additional property that the trolley *travels fast*. This constraint simply requires that the speed of the trolley is large enough, when the trolley is far from the goal position. Constraints on parameters for strategies S_{q2} are the same as for S_{q1} with additional constraint on p_2.

Note that the last two constraints (*smooth and fast at* X_{mid} and *travel fast*) are not necessary. If we do not use them, strategies are still successful in terms of reaching the goal position under the task limits, but some of them are too slow and fail to conclude the task in the required time of 180 seconds.

To investigate how successful those strategies are, we tested strategies $S_{q1}(X_{mid}, DX_{mid1}, k_1, p_1, p_2)$ with different parameters: 2 for k_1 and DX_{mid1}, 3 for X_{mid} and 5 for p_1 and p_2. Parameters were taken uniformly spaced from the parameter's range. These parameter settings give 300 different qualitative strategies from S_{q1}. 180 of these strategies obey additional constraint *travel fast* and belong to S_{q2}. All of these 180 different quantitative strategies are successful and very fast. The best time is 38 sec. (see Fig. 4), mean time 48 sec. and the worst time 78 sec. Considering the fact that the fastest operator's trace completes the task in 51 sec., the results are excellent. Out of the other 120 strategies, which do not respect the constraint *travel fast* and do not belong to S_{q2}, 100 are successful with mean time 84 sec. The details are given in Fig. 5.

Test of robustness

The strategies S_{q0}, were also tested for their robustness against changes in the start state, the test where the controllers in the form of decision or regression trees usually fail. In these experiments X_{mid} was set to values 15, 30, 45 and k_1 to $DX_{max}/4$. For each of these parameters settings, we generated 10 sets of random functions f_1^+, h_1^+, h_2^+, h_3^- with $R = 10$. In this way we get 30 quantitative strategies. All of those strategies were tested for their robustness against changes in the start state. Starting horizontal position was modified by 10% and 50%, and starting rope angle by -20%, 0% and 20%. This amounts to six combinations – six different tasks. Since controllers based on generalized operator's trajectory are goal directed, the change of the start state didn't affect their performance significantly. Out of $30 \times 6 = 180$ combinations, 162 were successful. The percentage of successful strategies with the changed start state is the same as the percentage of successful strategies with the original start state.

Discussion and conclusion

Our approach involves learning separately (a) the system's dynamics and (b) the generalized trajectory the operator is trying to follow. Once the trajectory is known, a controller can easily be constructed since we know the desired next state and a local model of system's dynamics near the current state. The induced control strategy often concludes the task faster than the original operator's traces. It is robust with respect to the changes in the system's dynamics, since the system's dynamics is learned apart from the trajectory. Qualitative strategy, generalized from the operator's trajectory, is comprehensible and offers the possibility to improve the operator's control strategy.

Some specific points of interest are:

1. To understand the operator's control strategy, it is important to identify the subgoals that the operator is pursuing at various times. A method for identifying discrete subgoals was developed in (Šuc & Bratko 1997). The generalized operator's trajectory, applied in this paper, can be viewed as a continuously changing subgoal and enables skill reconstruction in more difficult domains. Moreover, it opens new perspectives to the reconstruction of human control skill, such as qualitative treatment.

2. The induced strategy, in the form of the trajectory gives good insight in the operator's control strategy. Since it is expressed in a symbolic way, it is easy to understand, analyze and correct by considering the knowledge of the control task. One example of such analysis is given this paper. We used the fact that the rope length at the goal position X_g should be L_g to correct the learned L_{des} trajectory.

3. The generalized operator's trajectory enables the derivation of qualitative control rules from the induced control strategy. We believe that qualitative rules are closer to human thinking about control strategies, are easier to understand and provide better insight into what the operator is trying to do.

4. Given a qualitative control strategy, we can tune a controller within the corresponding qualitative constraints. In this way a qualitative control strategy defines an optimisation space for better performance. As shown in this paper, we can also use some background knowledge and derive constraints on the tuned parameters. By tuning (giving the qualitative parameters like X_{mid}, *increasing function*, real numerical values) we can achieve better performance than that of the source quantitative strategy.

5. By using a learned qualitative strategy and considering knowledge of the control task to find constraints on qualitative parameters, the learned strategy can be used for considerably different tasks (different goal position, different safety requirements, different parameters of the simulator). For example we could use the learned qualitative strategy to move the load (possibly different) backwards from the goal position to the start position with more rigorous safety requirements (for example $DX < 1\text{m/s}$ and $\phi < 6$ deg).

6. Regarding the derivation of a qualitative control strategy, one idea (Bratko 1997) is to induce qualitative rules directly from the execution trace. In this case a machine learning technique would be applied to a qualitative description of the execution trace. To use such an induced qualitative control strategy for controlling the system, the qualitative strategy would have to be operationalized by estimating numerical values of several parameters. The experimental results presented in this paper, and similar results in the Acrobot domain (Šuc & Bratko 1998), show the potential of such an approach. These ideas are the subject of our current research.

References

Bratko, I.; Urbančič, T.; and Sammut, C. 1998. Behavioural cloning of control skill. In Michalski, R.; Bratko, I.; and Kubat, M., eds., *Machine Learning and Data Mining: Methods and Applications*, 335–351. John Wiley & Sons, Ltd.

Bratko, I. 1997. Qualitative reconstruction of control skill. In Ironi, L., ed., *Qualitative reasoning : 11th International Workshop*, 41–52. Pubblicazioni N. 1036. Univ. of Pavia: Instituto di analisi numerica. Cortona (Italy), June 3-6.

Chambers, R., and Michie, D. 1969. Man-machine co-operation in a learning task. *Computer Graphics: Techniques and Applications* 179–186.

Cleveland, W. 1979. Robust locally-weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* 74:829–836.

Donaldson, P. 1964. Error decorrelation studies on a human operator performing a balancing task. *Med. Electron. Biol. Engng.* 2:393–410.

Forbus, K. 1984. Qualitative process theory. *Artificial Intelligence* 24:85–168.

Križman, V.; Džeroski, S.; and Kompare, B. 1995. Discovering dynamics from measured data. *Electrotechnical Review* 62:191–198.

Michie, D. 1993. Knowledge, learning and machine intelligence. In Sterling, L., ed., *Intelligent Systems*. New York: Plenum Press.

Sammut, C.; Hurst, S.; Kedzier, D.; and Michie, D. 1992. Learning to fly. In Sleeman, D., and Edwards, P., eds., *Proceedings of the Ninth International Workshop on Machine Learning*, 385–393. Morgan Kaufmann.

Schaal, S., and Atkeson, C. 1994. Assessing the quality of learned local models. In Cowan, J.; Tesauro, G.; and J., A., eds., *Advances in Neural Information Processing Systems*, volume 6. Morgan Kaufmann.

Urbančič, T., and Bratko, I. 1994. Reconstructing human skill with machine learning. In Cohn, A., ed., *Proceedings of the 11th European Conference on Artificial Intelligence*, 498–502. John Wiley & Sons, Ltd.

Šuc, D., and Bratko, I. 1997. Skill reconstruction as induction of LQ controllers with subgoals. In Mellish, C. S., ed., *Proceedings of the 15th International Joint Conference on Artificial Intelligence*, volume 2, 914–920. Nagoya, Japan, August 1997.

Šuc, D., and Bratko, I. 1998. Skill modelling through symbolic and qualitative reconstruction of operator's trajectories. Tech.Report, Faculty of Computer and Information Sciences, Artificial Intelligence Laboratory, University of Ljubljana, Slovenia.

Šuc, D., and Bratko, I. 1999. Symbolic and qualitative reconstruction of control skill. In De Raedt, L.; Muggleton, S.; and Michie, D., eds., *Machine Intelligence 16*. Springer Verlag. Forthcoming.