ОБНАРУЖЕНИЕ ЭМПИРИЧЕСКИХ ЗАКОНОМЕРНОСТЕЙ
(Вычислительные системы)
1999 год Выпуск 166

УДК 519.95

MACHINE LEARNING FROM HUMAN EXPERT PERFORMANCE OF DYNAMIC CONTROL TASKS[1]

Donald Michie and Jean Hayes Michie[2]

Abstract: Uses of stored skill-models to accelerate simulator-based real-time training in a control skill are discussed. A real-time coach must deliver advice at three levels: (1) what to do next, (2) what to watch for, and (3) what went wrong. Human learning and machine learning results are presented using different screen representations of a pole-and-cart balancing task.

1. INTRODUCTION

The first demonstration that a machine can learn a real–time control skill by imitating a human was made by Donaldson (1960). He used a mechanical apparatus in which a supporting body (hereafter called "the cart") was motor–driven back and forth along a track while supporting an inverted pendulum (hereafter called "the pole"). A skilled human sent signals to the motor such that the pole remained balanced and the cart remained within the confines of the track. The machine "watched" for a while, and then reproduced the skill.

---

[1]Машинное обучение динамическому управлению, основанное на подражании действиям эксперта

[2]AI Applications Institute, University of Edinburgh, UK E-mail: D.Michie@ed.ac.uk.

If humans were as capable of "observational learning" as Donaldson's machine, then the aircraft industry could save many million dollars annually. By simply watching skilled performance each trainee pilot would replace the hundreds of hours now spent in trial-and-error learning on costly simulators. This is of course a fantasy. Nevertheless, the means may now exist substantially to reduce training costs by integrating modern derivatives of the Donaldson feat into conventional simulator training. If a machine-learned model of a human skill can be embedded in a training simulator, then perhaps it can be got to perform on-line coaching functions, guiding the trial-and-error efforts of trainees. But the first step must be to understand how unaided humans progress through the learning regime, and what makes some trainees progress faster than others.

Thus our ultimate idea is to incorporate into the simulator's software an autopilot "cloned" from skilled human performance. The clone does not itself pilot the simulator but instead delivers real-time advice in the manner of an experienced flying instructor. For this to be possible, the clone's advice must come across to the trainee as intelligible and to the point. Therefore, unlike Donaldson's purely correlational formulation, it must be structured in terms of human-like rules, constraints and goals.

2. MOTIVATION AND APPROACH

After Donaldson's work, experimental studies of rule learning, both by imitation and also under trial and error conditions, have mainly been conducted with computer-based simulators. The motivation in the present work arises from the high cost of simulator-training in the aerospace industry. Some 25% of candidates entering UK helicopter courses fail to meet criterion. The attendant waste of money and people prompts the idea, sketched above, that the simulators themselves could be endowed with active tutorial intelligence. As a preliminary we have simulated a Donaldson-type laboratory control task for laboratory study. The aim is to examine the magnitude and causes of human responses to simulator training situations.

Donaldson's discovery was remarkable since complete sensorimotor control skills cannot be acquired by humans through any amount of passive watching. Would-be learners gain something from this

form of learning (termed by psychologists "observational learning"), and even more from well-judged tutorial advice. But the essential underpinnings of sensorimotor skill-acquisition require an irreducible minimum of sheer trial and error practice. The point is simply that the hundreds of "flying hours" consumed by trainee pilots on computer-simulated missions may greatly exceed this minimum.

Two ways of accelerating the expensive process suggest themselves. First, following each computer-detected blunder, a short period of passive watching could be interjected with a computer-generated replay with corrective commentary on what went wrong. Second, the trainee's trial and error practice could be accompanied in real time by a computer-generated voice-over on (1) what to do next, (2) what to watch for. Preliminary trials have convinced us that the uses of (1) are limited and that the rewarding gains lie in (2). Further analysis suggested that the role of (2) should be to promote the occurrence of the occasional "insights" reported by some of our subjects and reflected in their recorded traces. Insight in this sense has been defined as "a sudden reorganization or restructuring of the pattern or significance of events allowing one to grasp relationships relevant to the solution. Here insight represents a kind of learning and is characterized in an all-or-none fashion." (Reber, 1985). But before a stored skill-model can be used to deliver intelligible advice, it must itself be cast in intelligible form, in which the logic of decision is explicitly represented as rules and patterns.

3. RULE-BASED MODELS OF CONTROL

Chambers and Michie (1969) used computer simulation of Donaldson's pole-balancing task for bench testing their trial-and-error learning algorithm. In contrast with Donaldson's numerical and correlational representation, the learned models were essentially rule-structured. They then modified their algorithm for what we now call human-computer, or two-way, learning. Both agents learned in parallel while contributing to a merged skill-base. Chambers and Michie also looked at the same setup after disabling the feedback from the machine's own decisions. They thus obliged it to build its skill-base by cataloguing purely the outcomes of its human partner's situation-action behaviour, i.e. by imitation as in Donaldson. In

this behavioural cloning mode they noted what today is termed the "clean-up" effect. Once built, cloned skills commonly outperform the levels of the skilled humans from whom they are acquired.

A further feature of the Chambers-Michie study of the use of the simulator to record and analyse human trial-and-error learning, was later singled out for comment by Michie, Bain and Hayes Michie (1990):

"Experimental subjects were divided [by Chambers and Michie] into two groups. One group saw an animated picture on the screen. ...For the second group this was replaced by an animated image of four horizontal lines of fixed length, along each of which a pointer wandered back and forth. The subjects were unaware that the pointers corresponded to the current values of the four sensed state variables, cart-position, velocity, pole angle and angular velocity. ...Whenever any of the pointers ran off the end of its line in either direction, the FAIL message appeared and a new trial was initiated. In all other respects, the subjects in the two groups faced identical learning situations. They used a light-pen to administer LEFT and RIGHT decisions to the computer simulation, regardless of which of the two graphical representations was employed. The learning curves of the two groups were found to be indistinguishable. Thus Chambers and Michie had isolated a pure "seat-of-the-pants" skill, divorced from complications arising from the subjects' powers of cause-and-effect interpretation."

In introducing their repetition and extension of the Chambers and Michie imitation-learning results, Michie, Bain and Hayes Michie comment on the acquisition by humans of "capabilities which they cannot articulate", and they continue:

"... by a straightforward programming trick the models which are inductively inferred from such human-generated decision data can automatically be endowed with a self-explanation facility, and thus rendered articulate. The end-products can justly be described as articulate models of inarticulate 'subcognitive' skills."

This by-product of cloning may be termed "behavioural profiling". For the practical objective of real-time instruction during simulator-training, we do not see such profiling as a side-effect, but rather as the main deliverable. The plan, validated as feasible in

preliminary trials, is that a stored model, whether hand-crafted or cloned, should deliver tutorial advice to the trainee by voice-over.

4. LEARNING TO FLY

The next step in the cloning story was to see whether the findings could be scaled up to more complex real-time tasks. The piloting of simulated aircraft was chosen, as described by Sammut, Hurst, Kedzier and Michie (1992). The following summary of the flight simulator experiments is condensed from an account by D. Michie and C. Sammut (1995).

A flight simulator program is modified to log the actions taken by a human subject as he or she flies a simulated aircraft. The log file is used to create the inputs to an inductive learning program. The quality of the output from the induction program is tested by running the simulator in autopilot mode where the autopilot code is derived from the decision tree (equivalent to production rules) formed by induction. At the University of New South Wales (UNSW) source code to a flight simulator was made available by Silicon Graphics Inc., and the task was to fly a Cessna. In confirmatory studies by Camacho at the Turing Institute, continued at the Oxford University Computing Laboratory, the public-domain ACM flight simulator was used with the more difficult task posed by a simulated combat plane (Michie and Camacho 1994).

A feature again noted in both studies was that autopilots performed more consistently than the human exemplars. This effect of the induction process is explained by the fact that human behaviours which are not repeated in a consistent fashion are eliminated from the machine's digest as 'noisy data'. This clean-up effect is among the constant findings in cloning work that Urbancic and Bratko list in their 1994 review. They discuss pole-and-cart balancing, flight-simulator control, telephone-line scheduling and their own work on crane-simulator control, concluding as follows:

1. Successful clones have been induced using standard machine learning techniques in all four domains.

2. The clean-up effect, whereby the clone surpasses its original, has been observed in all four domains.

3. In all domains best clones were obtained when examples from a single human only were used.

4. The present approach lacks robustness in that it does not guarantee inducing with high probability a successful clone from given data.

5. Typically, the induced clones are not sufficiently robust with respect to changes in the control task.

6. Although 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.

The present work seeks remedial ground for the last-listed disability by first studying human learning behaviour. Using computer simulated variants of the original pole and cart control problem, attention is focussed on:

* the learning agent's causal concepts
* the learning agent's goal patterns.

5. HUMAN LEARNING OF THE LAB TASK

Past studies of perceptual and motor skills and their acquisition have generally been limited either to in-depth studies of individuals, or to group averages. Valuable conclusions have emerged, e.g. the "Law of Practice" relating speeds of task-execution to amounts of prior practice, Fitts' law relating speed of movement to distance and size of target, and various information-theoretic relations between diversity of choice and response times. But they have been limited by lack of data on variation among individuals given equal exposure to given practice regimes. Our experiments supplied answers to the following questions.

1. ***Question***: How large is individual variation of learning rates?

Answer : Variation turned out to be exceptionally large, with some subjects learning many times faster than others. This is in striking contrast with the general rule with biological traits. In groups of similar genetic and environmental background, standard deviations of psychophysical measurements tend to fall into the range 10in our real-time control task, coefficients of variation in raw scores commonly exceeded 100%. This striking phenomenon finds confirmation in extensive observations by Urbancic and Bratko

(1995) of human dynamical control of simulated container cranes. They state that: "remarkable individual differences were observed regarding the speed of learning as well as the speed of controlling, the frequency of successful experiments [trials] and the characteristics of the strategy used."

2. ***Question***: Are any measurable properties of subjects predictive of these large variations?

Answer: Gender, age and educational background were all clearly implicated, but in themselves only accounted for a part of the large inter-subject variability.

3. ***Question***: To what extent can the variations be attributed to experimentally varied conditions, including the task's running speed and its visual representation (animated cartoon versus moving indicators)?

Answer: At slow speeds unpractised subjects were aided by the cartoon representation, but subsequent learning rates under practice were not detectably affected by variation either of representation or of speed, alone or in combination.

4. ***Question***: To what extent do effects of these controlled variables throw light on the subjects' use of mental models, e.g. declarative versus procedural?

Answer: The finding under 3 above can be related to the longer times needed for deliberative interpretation of causal models than for reactive responses. Although pre-existing mental models can be of use under slow-running conditions in support of the beginner's early attempts, they play no detectable role in subsequent learning rates.

5. ***Question***: By separating out averaged learning curves into individual traces, can we say whether subjects learn in a more or less continuous fashion, or were there slow climbs occasionally punctuated by leaps?

Answer: In the upper performance quartile of learners, the second pattern was observed. Within the relatively limited total practice exposure of 200 minutes, the slower learners did not seem to show identifiable leaps. In consequence they remained so deficient in insight into the task's subgoal structure as severely to limit their ability to increase their skill.

6. ***Question***: Is this explainable in terms of previously described learning phenomena?

Answer: Earlier accounts of learning through intensive practice suggest that periods in which a procedural mental model is "tuned" are punctuated in some subjects by occasional "insights". For example Seashore (1951) writes: "The sudden progress associated with insight learning is probably to be attributed very largely to the perceiving of new qualitative patterns of action. On the other hand, the relatively steady progress frequently observed in learning of all types is usually attributable to refinements within a pattern."

7. ***Question***: Do those who learn faster, with "insight leaps", derive an associated improvement in explicit understanding of the task?

Answer: Scores on a questionnaire to measure explicit understanding correlated positively ($r > 0.5$) with final performance levels.

8. ***Question***: Is watching a task performed by another, as opposed to practising the task oneself, sufficient to impart explicit understanding?

Answer: A small group were allowed to watch the playback of samples of beginner performance and of expert performance. After watching a five-minute sample of beginner performance and five minutes of expert performance, questionnaire answers were comparable to those of subjects who had had 200 minutes of hands-on practice. But when they then attempted the control task for themselves, they performed like beginners.

6. PROTRACTED PRACTICE

There are no known limits to skill acquisition under continued practice, which in some cases has been followed over periods as long as 30 years. But a "diminishing returns" effect becomes conspicuous as less and less remains to be "automatised". To get a comparative idea of these more intensively practised states the performance of a subject was investigated with approximately 100 times more practice on the pole-and-cart task than the main sample (more than 300 hours spread over a number of years). To each of the questions: "Is learning still continuing?" and "Do new insight leaps still occur?", the answer was positive.

7. DISCUSSION

If insight leaps correspond to acquisition of new subgoals, far-reaching implications follow for the design of the proposed real-time advice modules. Unstructured induction of "flat" skill-models from recorded traces of expert behaviour should be relinquished in favour of structured induction in the style of Shapiro (1987). The latter bases itself on the structuring of a skill in the form of a hierarchy of subgoals. Each of these will be interpretable in the present context as a spoken "what to watch for" alert in appropriate screen-displayed situations coupled with an associated "What to try for", — thus "Look for the centre approach moment — now try for a reversal of the pole's lean."

The reader may wonder whether advice giving skill models have necessarily to be machine-learned. Could they not better be hand-crafted? Control theory is not at present equipped with methods for generating goal-hierarchies and situation-action rules. But an expert performer may seek to express control skill in a machine-interpretable production-rule language that can handle these. Preliminary experience suggests that for a task as simple as the pole and cart, an expert can write a production-rule control program that simulates the skill of an advanced trainee, but falls far short of his own logged behaviour. Inductive extraction from the latter of machine-executable skill models fares better. We conclude that, especially with more complex elaborations of such tasks, behavioural cloning will turn out to be an indispensable tool in constructing the automated computer tutors of the future.

REFERENCES

Donaldson, P.E.K. (1960). Error decorrelation: a technique for matching a class of functions. Proc. III Internat. Conf. on Medical Electronics, pp. 173-178.

Chambers, R.A. and Michie, D. (1969) Man-machine co-operation in a learning task. In Computer Graphics: Techniques and Applications (eds. R. Parslow, R. Prowse and R. Elliott Green). London: Plenum Publishing Co., pp. 179-186.

Michie, D., Bain, M. and Michie, J.E. (1990) Cognitive models from subcognitive skills. In Knowledge-Based Systems in Industrial

Control (eds. M. Grimble, J. NcGhee and P. Mowforth), Stevenage, UK: Peter Peregrinus, pp. 71–99.

Michie, D. and Sammut, C. (1995) Behavioural clones and cognitive skill models. In Machine Intelligence 14, Oxford University Press, pp. 387–395.

Reber, A.S. (1995) The Penguin Dictionary of Psychology, Penguin Books.

Sammut, C. Hurst, S, Kedzier, D. and Michie, D. (1992) Learning to fly. In Proc. 9th Internat. Machine Learning Conf (ML92) (ed. D.H. Sleeman), San Mateo, CA: Morgan Kaufmann.

Seashore, R.M. (1965) Work and motor performance. In Handbook of Experimental Psychology, 7th edition (ed. S.S. Stevens), John Wiley, pp. 1341–62

Shapiro, A.D. (1987) Structured Induction in Expert Systems, Addison–Wesley.

Urbancic, T., and Bratko, I. (1994). Reconstructing human skill with machine learning. In Proc. Europ..Conf..on AI (ECAI 94), Amsterdam.

Urbancic, T., Bratko, I. (1995) Controlling Container Cranes: A Case-Study in Reconstruction of Human Skill. Electrotechnical Reviews, 62 (3-4), Ljubljana, 199-205. (also in AIT'95 Proceedings (Artificial Intelligence Techniques), eds. J.Zizka, P.Brazdil, Brno, 1995, 113–127)