OPTIMIST: A New Conflict Resolution and Learning Algorithm

Roman V. Belavkin (r.belavkin@mdx.ac.uk)

School of Computing Science,
Middlesex University, London NW4 4BT, UK [a]

June 9, 2003

[a]This work is sponsored by ESRC Credit and the ORS Award Scheme

OVERVIEW

Optimism + Optimisation = Optimist

- Motivation for this work
- Some established conflict resolution methods
- Method description
  - Recursive estimation of expected cost
  - Conflict resolution
- Demo application (search space)
- Method performance and properties

COGNITIVE MODELLING

Unified Theories of Cognition

(Newell, 1990)

$\Downarrow$

SOAR | ACT–R | $\cdots$ | Neural Nets

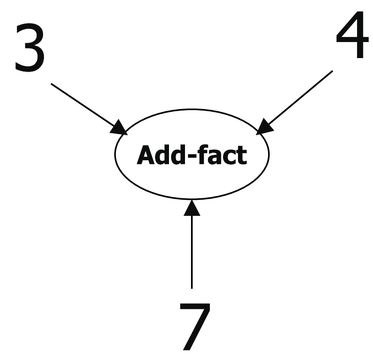
$\Downarrow$

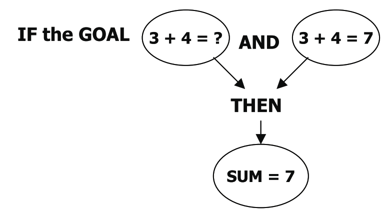
Cognitive Models + Data

ACT–R COGNITIVE ARCHITECTURE

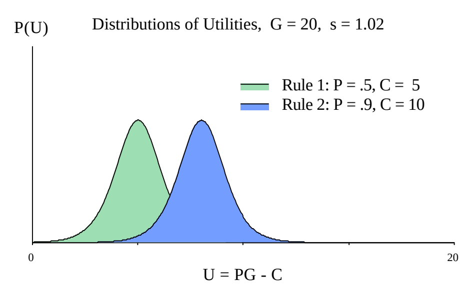
(Anderson & Lebiere, 1998)

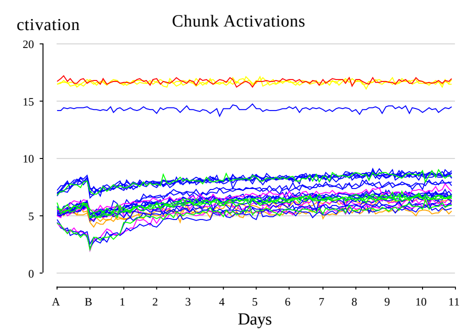
Symbolic level: Facts, goals (declarative knowledge), rules (procedural knowledge), goal–stack, perception–action buffers





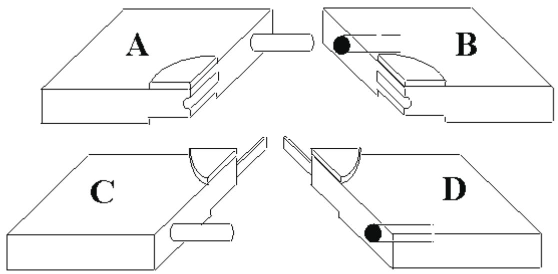
Sub–Symbolic level: Activations, association strengths, utilities, probabilities, time decay, etc



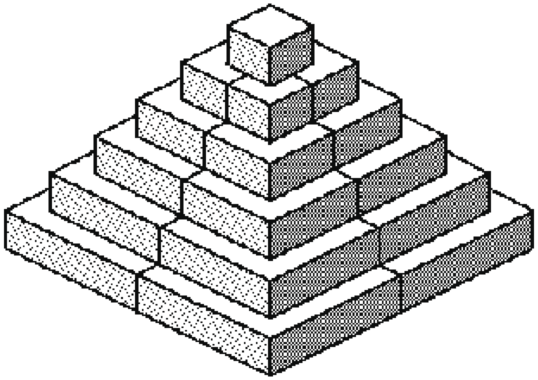


CONFLICT RESOLUTION

Usually there are many ways to go from the initial (current) state to the goal state:



$\longrightarrow$



- Traditional conflict resolution strategies: refraction, recency, specificity, priority, etc.
- In effect conflict resolution strategy implements particular search method
- In ACT–R conflict resolution is also a model of choice behaviour and decision–making

CONFLICT RESOLUTION IN ACT–R

In ACT–R (Anderson & Lebiere, 1998) each alternative i is represented by a production rule in a conflict set. A rule that wins should have the highest *utility*:

$$\boxed{U_i = P_i G - C_i + \text{noise}(s)}$$

rule's properties :

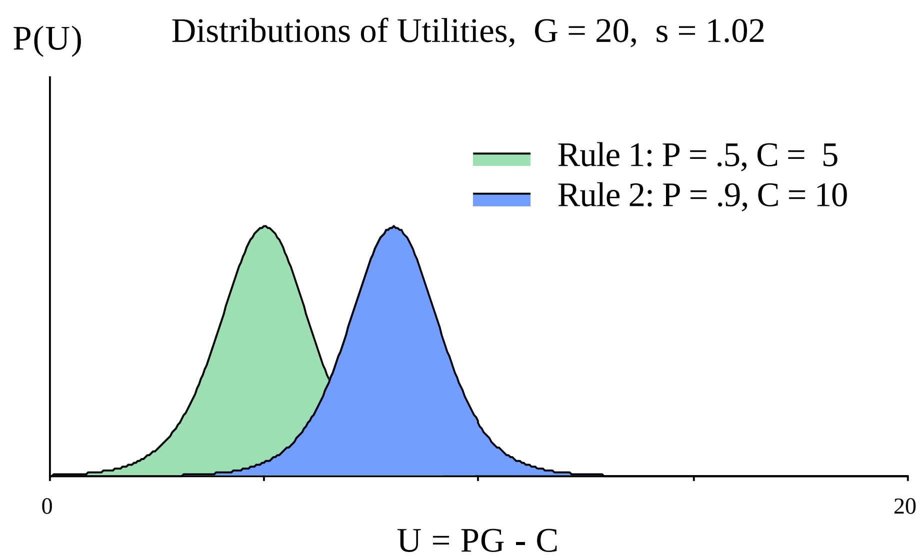
P_i — probability

C_i — cost (e.g. time)

global parameters :

G — goal value (in time units)

s — controls the noise variance



COST & PROBABILITY

Let C be a random cost of achieving the goal, and let $P(C)$ be the probability to achieve the goal at cost C. The expected value of the random cost is

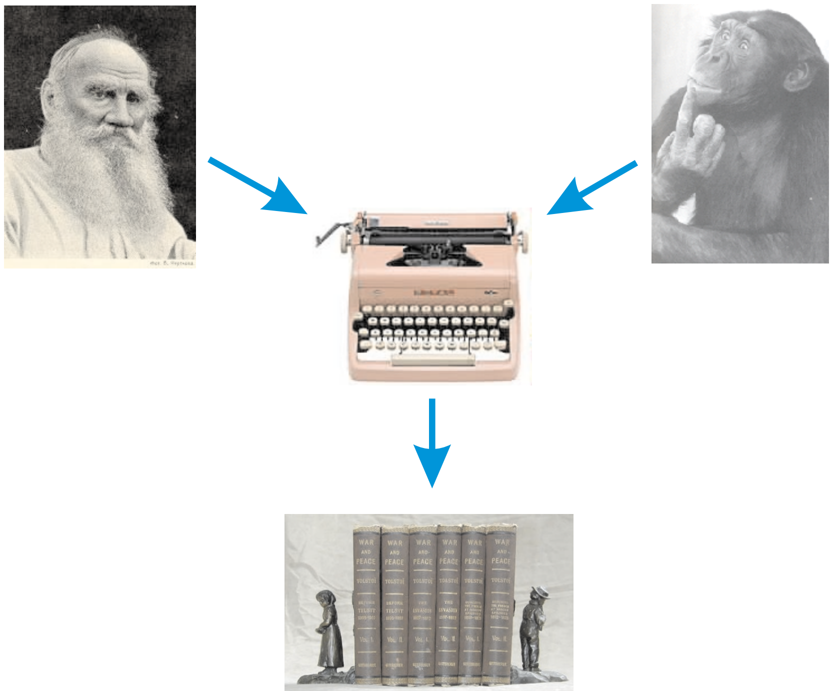
$$E\{C\} = \sum_C C\,P(C) \quad \left(E\{t\} = \int_0^\infty t\,P(t)\,\mathrm{d}t \right)$$

The conflict could be resolved by choosing rule i:

$$i = \arg\,\min E\{C_i\}\ .$$

Unfortunately, we do not know the distributions $P(t)$.

LEO TOLSTOY vs A CHIMP



- It took Leo Tolstoy 7 years to write “War & Peace”.
- A chimp can *probably* type it in $\sim 10^{10}$ years.
- How long one should wait before giving up?

$$E\{C\} \leq G < \infty$$

GOAL STATE AS A POISSON PROCESS

$$\boxed{P(n \mid \lambda) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}} \quad (1)$$

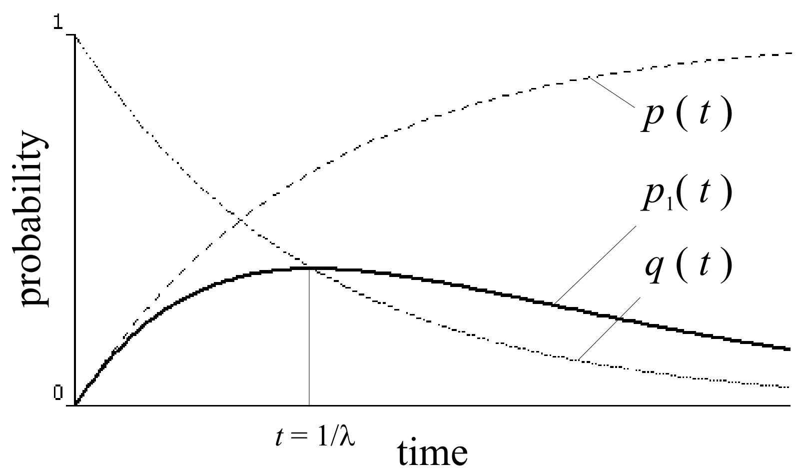
where λ is the *mean count rate* ($1/E\{C\}$), and $n = 0, 1, 2, \ldots$ is the number of successes.

If the goal is possible, then $E\{C\} < \infty$ (or $\lambda > 0$).

$q(t)$ — prob. of failure ($n = 0$)

$p(t)$ — prob. of success ($n > 0$)

$p_1(t)$ — prob. of 1st success ($n = 1$)



ESTIMATION OF THE EXPECTED COST

The rate λ is unknown, but we know t amount of time (efforts) spent and n number of successes. To estimate λ (and, hence, $E\{C\} = 1/\lambda$) we need posterior distribution

$$P(\lambda \mid n) = \frac{P(n \mid \lambda)P(\lambda)}{P(n)}$$

If we assume that all values of λ are equally probable, then

$$P(\lambda \mid n) = tP(n \mid \lambda) \; .$$

Indeed, it is possible to show that above is true for $P_\varepsilon(\lambda) = \varepsilon e^{-\varepsilon\lambda}$ when $\varepsilon \to 0$.

POSTERIOR ESTIMATION

- Maximum likelihood:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}P(\lambda \mid n) = 0 \quad \Rightarrow \quad \lambda = \frac{n}{t} \quad \Rightarrow \quad E\{C\} = \frac{t}{n}$$

- Maximum of posterior estimate:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}tP(n \mid \lambda) = 0 \quad \Rightarrow \quad \lambda = \frac{n}{t} \quad \Rightarrow \quad E\{C\} = \frac{t}{n}$$

- Posterior mean estimate:

$$E\{\lambda\} = \int_0^\infty \lambda P(\lambda \mid n)\mathrm{d}\lambda = \frac{n+1}{t} \quad \Rightarrow \quad E\{C\} = \frac{t}{n+1}$$

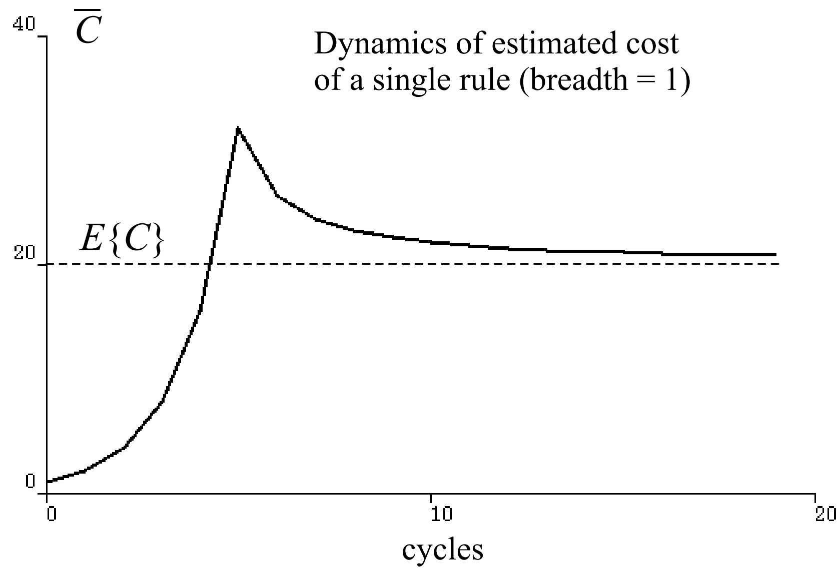
RECURSIVE ESTIMATION

Let us restart the computer (give up) after Δt and see if it reaches the goal state (binomial trial).

Starting with some $\Delta t_0 = C_{\min}$ (e.g. ≈ 50ms) let us set each next Δt to last *estimated cost*

$$\Delta t_{k+1} = \bar{C}_k = \frac{t_k}{n_k + 1}$$

$$\lim_{k \to \infty} \bar{C}_k = E\{C\}$$



THE CONFLICT

For each option x let us record $t(x)$ — efforts, $n(x)$ — n. of successes, $k(x)$ — n. of times x has been used. To resolve the conflict we introduce a *random prediction* C_ξ:

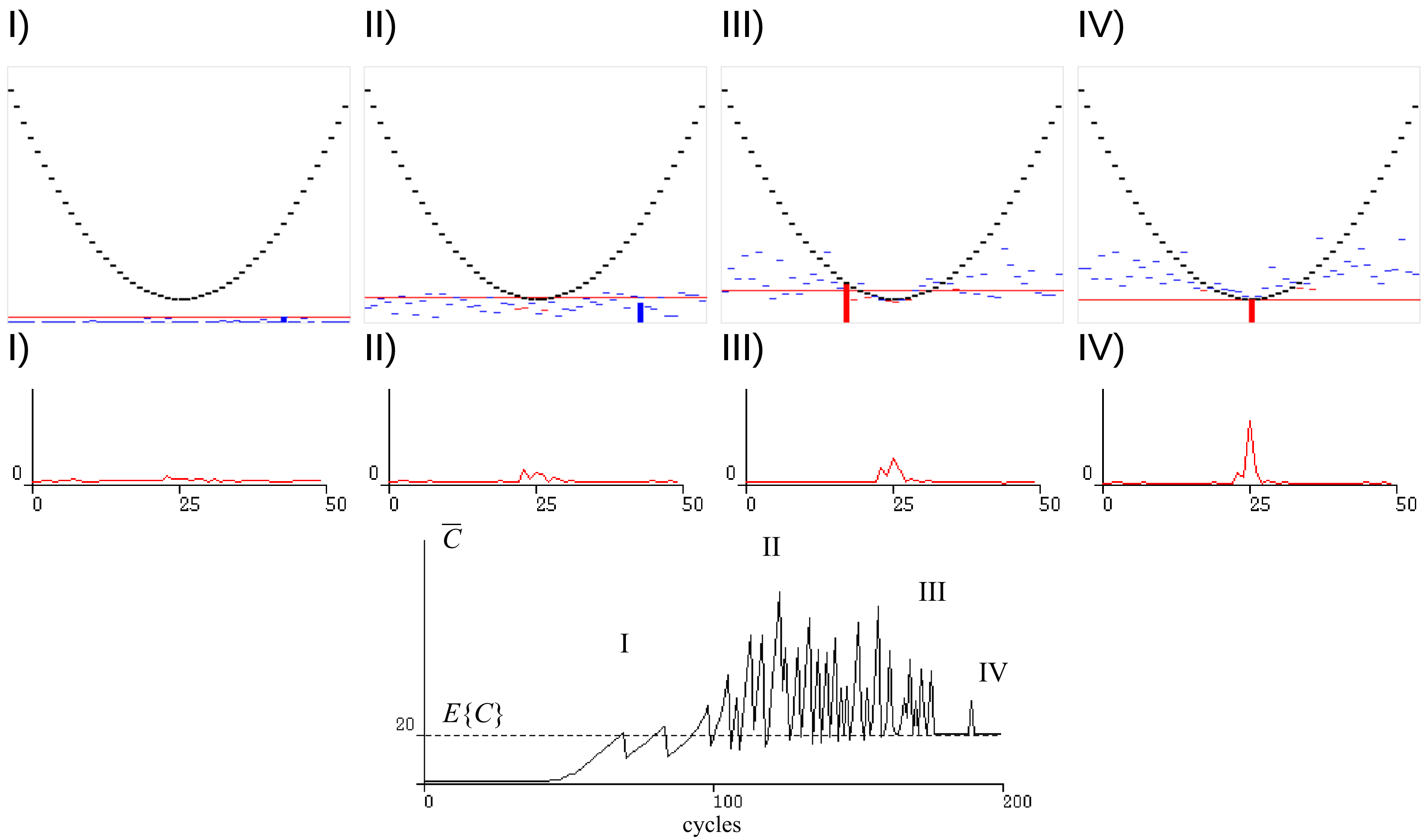
$$E\{C_\xi(x)\} = \bar{C}_{k(x)}(x) \qquad \text{e.g. } C_\xi = \mathrm{rand} \in [0, 2\bar{C}]$$

The *random estimated cost* is defined as

$$\tilde{C}(x) = \frac{k(x)\bar{C}(x) + C_\xi(x)}{k(x) + 1}$$

Conflict resolution: $\qquad x = \arg\,\min \tilde{C}(x).$

BEHAVIOUR OF THE ALGORITHM



IMPORTANT PROPERTIES

- The method possess all the properties of the current ACT–R mechanism plus dynamic goal value (i.e. $\bar{C}$) and noise variance (i.e. random prediction).
- Noise is rule–specific and its effect is a function of experience.
- The use of Poisson distribution is supported by some studies on animal choice behaviour and learning (Myerson & Miezin, 1980; Mark & Gallistel, 1994).
- The plasticity effect is present in biological neurons (Sejnowski, 1977a, 1977b).
- Breadth and depth of the search are adaptive and controlled automatically.
- The algorithm behaviour is in some way similar to 'simulated annealing' (Kirkpatrick, Gelatt, & Vecchi, 1983) or 'taboo' search (Glover, 1977). However, its performance compared to these methods is yet to be established.

QUESTIONS & FUTURE WORK

- How to implement different kinds and degrees of reinforcement (reward, penalty)?
- How can the method be extended to incorporate multi–objective optimisation (not only time)?
- How does the algorithm compare to other search techniques?

QUESTIONS?

References

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Lawrence Erlbaum Associates.

Glover, F. (1977). Heuristics for integer programming using surrogate constraints. *Decision Sciences*, *8*, 156–166.

Kirkpatrick, S., Gelatt, C. D., & Vecchi, J. M. P. (1983). Optimization by simulated annealing. *Science*, *220*(4598), 671–680.

Mark, T. A., & Gallistel, C. R. (1994). Kinetics of matching. *Journal of Experimental Psychology*, *20*(1), 79–95.

Myerson, J., & Miezin, F. M. (1980). The kinetics of choice: An operant systems analysis. *Psychological Review*, *87*(2), 160–174.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, Massachusetts: Harvard University Press.

Sejnowski, T. J. (1977a). Statistical constraints on synaptic plasticity. *Journal of Mathematical Biology*, *69*, 385–389.

Sejnowski, T. J. (1977b). Storing covariance with nonlinearly interacting neurons. *Journal of Mathematical Biology*, *4*, 303–321.