The particle swarm: Social adaptation of knowledge

James Kennedy
Bureau of Labor Statistics
Washington, DC
Kennedy_Jim@bls.gov

Particle swarm adaptation is an optimization paradigm that simulates the ability of human societies to process knowledge. The algorithm models the exploration of a problem space by a population of individuals; individuals' successes influence their searches and those of their peers. The algorithm is relevant to cognition, in particular the representation of schematic knowledge in neural networks. Particle swarm optimization successfully optimizes network weights, simulating the adaptive sharing of representations among social collaborators. This paper introduces the algorithm, begins to develop a social-science context for it, and explores some aspects of its functioning.

Interpersonal information processing

Information processing of verbal symbols has frequently been described as a process which occurs inside peoples' heads. But note that it is impossible to dissociate the symbol-processing function of language from its communication function. Further, concept learning is rarely an independent activity, but more frequently occurs interpersonally, whether through literature, pedantry, informal conversation, or some other means (Markus & Zajonc, 1985). Finally, we observe that individuals are able to communicate with one another about their thinking; importantly, humans have many tools for expressing the methods they use to process information, and are capable of adopting the information-processing techniques used by their peers. The particle swarm paradigm is associated with those theories which describe psychological reality as a social construction. Further, this paper asserts that the method by which societies construct reality and operate upon it comprises an excellent optimization technique when applied to artificial systems; the particle swarm algorithm is useful as a mathematical tool as well as a social theory.

Multivariate trajectories

The algorithm assumes a multidimensional psychological space, however that is specified (i.e., a neural network or symbolic model); it is further assumed that some measure of goodness is possible. The nature of the dimensionality and of the goodness measure depend again upon a theory, and the particle swarm model is metatheoretical with respect to cognition, though a connectionist example will be used here to demonstrate the paradigm. Thus, in the current example the psychological space is a hyperspace of connections among nodes, and an individual's psychological position at any moment is described as a particle whose coordinates are connection weights. The vectors of weights for each individual i on dimension d will be called X_{id}.

Individuals conduct repeated experiments, modeled as an iterative loop in a simulation, each passing data through a network of weights X_{id} which are initially random. The outcome of each experiment is a measure of error, or the distance between calculated outputs and desired outputs as it is usually calculated in feedforward networks. At each iteration the individual's position in hyperspace is changed by adding a ΔX_{id} vector, called V_{id}, to the vector of coordinates. Thus, iteratively,

$$X_{id} \rightarrow X_{id} + V_{id}.$$

Unaltered, the ΔX_{id} vector would move individuals in straight lines toward infinity. The particle swarm algorithm optimizes by adjusting the amount of change.

The individual best

Each individual enters the coordinates of its current position into the formulas for the neural network and measures the error of the estimate of target values, then moves to a new position and repeats. As individuals move through the multivariate space, they compare their current error value to the best they have attained at any point up to that iteration. The best, that is lowest, error term encountered thus far is called $pbest_i$, and the position where that evaluation was attained is represented as the vector P_{id}. A feature of this paradigm which distinguishes it from other evolutionary computation paradigms is its reliance on the individual's memory of $pbest_i$ and P_{id}.

0-7803-3949-5/97/$10.00 ©1997 IEEE

If the present position is better than any previously encountered, then P_{id} is updated with the current weight coordinates. The difference $P_{id} - X_{id}$ indicates the distance between the individual's previous best and current positions. Each individual i then is associated with a "nostalgia" vector (Kennedy and Eberhart, 1995) of $P_{id} - X_{id}$ differences.

Each element in the vector of differences is weighted by a positive random number φ, whose upper limit is a parameter of the system. This new vector is added to the change vector, so that

$$V_{id} \rightarrow V_{id} + \varphi_{id}(P_{id} - X_{id})$$

Adding this new vector to the current coordinates X_{id} introduces a stochastic tendency to return toward the individual's previous best position.

Psychologically it is a tendency to remember and return to regions in the psychological space which have demonstrated promise, that is, to combinations of beliefs that have seemed good or reinforcing.

The neighborhood best

In order to add a social component, a neighborhood is defined for each individual. Though there can be endless ways to specify it, the present simulations define the neighborhood as comprising the target individual and the two individuals immediately adjacent in the array. Thus for each individual i the neighborhood is composed of individuals i-1, i, and i+1. This is equivalent to the "circle" communication pattern, as described in Leavitt (1951).

All individuals search the problem space simultaneously; each has its own pbest_i and P_{id}. For the social term of the algorithm, individual i determines which member of the neighborhood has achieved the best pbest_i so far. The index of the neighbor who has found the best position, that is, the one with the lowest error value, is assigned to the variable g, and that individual's best position is referred to as P_{gd}.

Thus the vector $(P_{id} - X_{gd})$ represents the distance from individual i's current position to the best position that has been found by any member of the neighborhood. The particle swarm algorithm weights every element of this vector, as before, with a random number φ defined by an upper limit, and adds this vector of social influence to the change vector V_{id}.

In sum, the particle swarm algorithm iterates through the following formula:

$$V_{id} \rightarrow V_{id} + \varphi_{id1}(P_{id} - X_{id}) + \varphi_{id2}(P_{gd} - X_{id})$$

with $X_{id} \rightarrow X_{id} + V_{id}$, for individual i on dimension d. The only additional rule is that V_{id} is limited by some value V_{max}. This limit serves three purposes: it keeps the computer from overflowing, it realistically simulates the incremental changes of human learning and attitude change, and it determines the granularity of search of the problem space. The limit is implemented simply, if not elegantly, by specifying that if $V_{id} > V_{max}$, then $V_{id} = V_{max}$, or if $V_{id} < -V_{max}$ then $V_{id} = -V_{max}$.

Comparison to backpropagation of error

The theoretical view of learning as backpropagation of error differs from the particle swarm view in two important ways (see Rumelhart, Hinton, and Williams, 1986). First, backpropagation implies that the individual is a kind of sealed container with information processing going on inside it. The particle swarm shifts the locus of information processing, and of mind itself, out into interpersonal space, with individuals discovering collaboratively how to process information. In other words, a society is viewed as a parallel distributed processing system comprising individual cognitive networks which adapt collaboratively. Secondly, while backpropagation of error is a deterministic hill-climbing algorithm, the particle swarm exploits the impartiality of randomness in the search of the problem space; individuals in the population are pulled toward optimal positions, but do not climb hills in or between optimal regions.

Importantly, the dynamical nature of the algorithm represents a departure from rational models of cognition. Whereas backpropagation of error depicts an individual reasonably changing his or her beliefs in order to make them more consistent with the facts as they are perceived, the particle swarm simulates individuals changing their beliefs in order to be more like their neighbors. Thus it is a *social*-psychological model of knowledge management.

Results of simulations

The XOR problem requires a network to map inputs of (1, 0) or (0, 1) to an output value of 1, and to map inputs of (0, 0) or (1, 1) to an output of 0. As the logistic function used in the network approaches limits of 1.0 and 0.0, it is common to replace target outputs with the attainable values of 0.9 and 0.1. The network's learning objective then is discover a set of weights which will accurately produce 0.9 when given two inputs which differ from one another, and 0.1 when the two inputs are identical.

A network was defined with 2 hidden nodes (the minimum required for this mapping, unless direct links from inputs to outputs are included in the model), and was trained until some member of the simulated population achieved an average squared error per node value less than 0.02. Thus each particle in the population moved in 9-dimensional space, in other words each individual needed to optimize a combination of nine floating-point numbers.

Particle swarm trials, varying V_{max} and φ

The simulation of the full particle swarm model was run, varying values for the two parameters V_{max} and φ, and observing how these values affected the efficiency of the algorithm, measured in terms of the number of iterations required for some member of the population to reach a criterion of average squared error per node < 0.02. Iterations are displayed as medians of 20 trials per cell, in a model with a population of 20 individuals.

Local optima The system was defined as trapped in a local optimum when it iterated more than 3,000 times without reaching the criterion. Local optimum results were tabulated but are only summarized here, in the interest of brevity. Some parameter combinations were rather vulnerable to local optima, and others were not. In general, when V_{max} was low, particles had more difficulty escaping from locally optimal regions. V_{max} determines how large steps through the data space each particle is allowed to take; when these steps are constrained to be small, individuals may be unable to step out of poor regions. As seen in Table 1, the particle swarm found appropriate weights in a relatively few interations.

Table 1. Median numbers of iterations required for a member of the population to attain an error criterion of e < 0.02, using the full model.

V_{max} \ *Psi*	0.5	1	2	4	6	8	10
0.5	150.5	121.5	139.5	130.5	110	124	145
1	157	95.5	88	91	84.5	106.5	102.5
2	134	127.5	99.5	71	84.5	84.5	84.5
4	143.5	101	137.5	82.5	62	79	90
6	131.5	117	107.5	103	86.5	73	98.5
8	122	104.5	93.5	131	118.5	92.5	78.5
10	72.5	101	121	97.5	112	79.5	104.5

The "cognition-only" model

The current paper alleges that cognitive science has tended to treat individuals as if they were isolates, and as if cognition occurred inside the head, privately. It is noted that one of the terms of the particle swarm velocity equation, $\varphi_{id}\,(P_{id} - X_{id})$, represents private thinking. Thus it was possible to test the cognitive part of the particle swarm algorithm by optimizing a neural net using the following velocity formula:

$$V_{id} \rightarrow V_{id} + \varphi_{id}\,(P_{id} - X_{id})$$

As seen in Table 2, this version of the adaptive algorithm was only slightly more vulnerable to failure than was the full model. It appeared that most of the time this version failed to converge within 3,000 iterations, the problem was more one of failure to find an optimal region than of being captured by local attractors. Individuals in this version tended to search the areas in which they had been initialized, and, at least when V_{max} and φ were both small, they failed to move into optimal regions. In median comparisons, the cognition-only model required more iterations than the full model in 47 of the 49 conditions tested.

Table 2. Median numbers of iterations required for a member of the population to attain an error criterion of e < 0.02, using the cognition-only model. Lemniscates indicate that more than half the trials exceeded 3,000 iterations.

V_{max} \ *Psi*	0.5	1	2	4	6	8	10
0.5	∞	∞	835	209	172	198.5	156.5
1	1261	878.5	319	170.5	141	122.5	115.5
2	490.5	354	225.5	134	120	111.5	127.5
4	231	257	185	136.5	137	104.5	102.5
6	218.5	165.5	169	141	132.5	99.5	99
8	192.5	166	145	133	115.5	113	102
10	146	129.5	132	144.5	132	121	90.5

In sum, the cognition-only model performed rather well. As an adaptive method, individual approximation to the optimum seems to function satisfactorily, though not nearly as well as the full model. This version was susceptible to failure only when both parameters were set very low, but required more iterations to satisfy the criterion.

The "social-only" model

Seeing that the "cognition-only" model worked rather well, one wonders how a "social-only" version of the algorithm would perform. Thus a model comprising the following formula was implemented:

$$V_{id} \rightarrow V_{id} + \varphi_{id}\,(P_{gd} - X_{id})$$

This model implies a social-psychological process with no special tendency for individuals to return to beliefs that had proven successful for themselves in the past. Rather, individuals compare the effectiveness of beliefs of neighborhood members and change toward those that are relatively successful. Note that a minimum of one individual in the population and a maximum of half of them can be the best performers in their own neighborhoods, and exert influence on themselves.

As seen in Table 3, the social-only algorithm converged faster than the full model in 47 of the 49 cells, and faster than the cognition-only version in all 49 cells.

Table 3. Median numbers of iterations required for a member of the population to attain an error criterion of $e < 0.02$, using the social-only model.

V_{max} \ Psi	0.5	1	2	4	6	8	10
0.5	93.5	95	88.5	98.5	96.5	93	102.5
1	89.5	72	66.5	65.5	70.5	72	69
2	83	62.5	57.5	64.5	52.5	53.5	53
4	86.5	61.5	72.5	48	60.5	54.5	57
6	67.5	95.5	76.5	90.5	77	71	74.5
8	110	87.5	75	76.5	61	76	34
10	115	106.5	70	63	78.5	69	81

Thus, the social algorithm is seen to be a more efficient optimizer for the present problem than both the cognition-only and the full versions of the particle swarm algorithm. Despite a slight susceptibility to be captured by local optima, the social version appears to be the best of the three models tested thus far, at least for this particular neural network problem.

The "selfless" model

As was noted above, in the social-only version individuals could be attracted to their own best positions when theirs was the best in the neighborhood. That is, as they compared performances of the three members of their neighborhoods, they included themselves. Thus, in some cases, where $i = g$, there is a confounding between self-influence and other-influence.

In order to eliminate the confounding effect, a "selfless" model was tested. This version was identical to the social-only version, with the exception that the neighborhood did not contain the individual's own previous best performance, that is, $i \neq g$. Thus none were attracted to their own successes, but rather only followed one another through the hyperspace.

As seen in Table 4, the selfless version met the criterion faster than the social-only version 20.5 times (a tie is rated as 0.5) out of 49. Thus, it seems to be slightly less efficient than the social-only model.

Table 4. Median numbers of iterations required for a member of the population to attain an error criterion of $e < 0.02$, using the "selfless" model.

V_{max} \ Psi	0.5	1	2	4	6	8	10
0.5	112	94.5	98	91	114	102	92
1	95.5	73.5	70	52.5	81.5	65	70.5
2	79	67.5	116	53	62.5	50	59
4	109	64.5	51.5	61	48	47.5	59
6	79	71	73	58	77	82.5	58
8	85.5	65.5	90	82	78	77	68
10	88	124	53	84	82.5	72	64.5

With 49 cells per version, matching cells from the four versions were paired to determine whether one model reached criterion in a significantly smaller number of iterations than the other. The normal approximation of the binomial sign test was used to assess the significance of differences. The null hypothesis is that, if two models are equivalent, about the same number of cells will require more and less iterations to converge; the sign test estimates the probability that the degree of inequality observed in the data would have been produced by chance. The full model converged faster than the cognition-only version, $z = 6.429$, $p < 0.001$, slower than the social-only model, $z = 6.429$, and slower than the selfless model, $z = 5.857$, $p < 0.001$. The social-only model converged faster than the cognition-only model in every cell, $z = 6.999$, $p < 0.001$, and faster than the selfless version, $z = 1.286$, not significantly. Finally, the selfless model was more efficient than the cognition-only model in all 49 cells, $z = 6.999$, $p < 0.001$.

Cellwise comparison may not be the ideal test of performance of these algorithms, especially if the optimal parameter combinations differ between models. The lowest median iterations in any cell for the various models were as follows:

Full model	62
Cognition-only	90.5
Social-only	34
Selfless	47.5

Thus this measure of performance parallels the previous one.

Discussion

First of all, note emphatically that these results represent performance of versions of the algorithm on one particular-- and very simple-- problem. Versions of the particle swarm algorithm will almost certainly perform differently on problems featuring higher dimensionality, greater nonlinearity, more local optima, etc.

The particle swarm paradigm is metatheoretical with respect to cognition. The present paper applies the optimization paradigm to the simulation of a feedforward neural net, but other successful applications include fuzzy cognitive maps (Kosko, 1992), parallel constraint satisfaction networks, and quantitative balance networks (Kennedy, 1995), besides various testbed functions (Kennedy and Eberhart, 1995).

The term "social cognition" has been established to refer to the special qualities of thoughts about social objects, with the special characteristics that these have, relative to other objects. In the particle swarm paradigm, social cognition is taken to refer to a kind of *cognition that is literally social*, i.e., the cognitive process itself is social.

In all versions of the present test it appeared that failure was associated with too-small values of V_{max}. With high V_{max}, individuals often speed past the target region, and discover even better positions than they set out for; further, they are able to take steps sufficiently large to escape from local optima. With V_{max} low, however, improvement can only arise from exploration around familiar regions. In order for the adaptive algorithm to work, individuals must be capable of escaping from regions which are minimally satisfying, to "change their minds."

The results from the cognition-only condition were surprisingly good. Whereas it had been anticipated that uncoordinated search would tend to result in failure (it does on other problems), this was the case in only two cells of the research matrix, though it should be noted that "success" here means only that one of the 20 individuals in the trial was able to meet the error criterion. Because they did not interact, the population of 20 was identical to running 20 individual trials.

The two social versions resulted in an even pleasanter surprise, easily outperforming the full model and cognitive-only conditions on this particular problem. Individuals in the social conditions accomplished the learning task by evaluating who in their immediate neighborhood had found the best solution so far, and then emulating that individual.

The so-called selfless version of the algorithm performed nearly as well as the social-only version, and better than the other two versions on this specific task. In the selfless model, individuals merely emulated one another, never being influenced by their own past findings. The selfless model is not theoretically justified, though it is interesting to note how much better it performed than the individualistic, cognition-only model.

In sum

The "cognitive" term of the formula can be easily thought of as an interpretation of Thorndike's (1911) "Law of Effect" (or Hull's, cf. 1930, "habit strength"), which stated simply that a random behavior which is followed by a reinforcement becomes more probable in the future. In the present case, the "behavior" is cognitive, which of course violates the positivistic ethos of behaviorism, but, more interestingly, this interpretation assumes that the attainment of correct or self-consistent knowledge is reinforcing. A model such as the present one presumes that the individual is motivated to reduce error. This is not distinguishable from Festinger's (1957) theoretical statement that cognitive dissonance is an aversive state equivalent to a "drive" in behaviorism. Dissonance was said to arise between two cognitive elements when the obverse of one followed from the other. The cognitive term of the particle swarm is shown to simulate venerable concepts from the classic era of the science of behavior.

The "social" half of the formula recalls Bandura's (1965; 1986) concept of modeling or vicarious reinforcement ("no-trial learning"). According to that theoretical perspective, observation of a model being reinforced for performing a behavior will increase the probability of the observer performing the behavior. Since we have already admitted that consistent or consonant thinking might be reinforcing, it follows that observation of a model who apparently is experiencing correct, valid, or internally consistent cognition should result in the observer's imitation of that cognition.

The psychological assumptions of particle swarm theory are general and noncontroversial: in their search for consistent cognitions, individuals will tend to retain their own best beliefs, and will also consider the beliefs of their colleagues. Adaptive change results when individuals perceive that others' beliefs are better than their own. The concepts are not new. What is new is the evidence from computer simulations that these simple concepts, taken together, create an information-processing technique which may be powerful enough to manage the huge amount of information comprising human knowledge.

References

Bandura, A. (1965). Vicarious processes: A case of no-trial learning. In L. Berkowitz (Ed.), *Advances in experimental social psychology, Vol. 2*, pp. 1-55.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* New Jersey: Prentice-Hall, Inc.

Festinger, L. (1957). A theory of cognitive dissonance. Evanston, Ill.: Row, Peterson.

Leavitt, H. J. (1951). Some effects of certain communication patterns on group performance. *Journal of Abnormal and Social Psychology, 46*, 38-50.

Hull, C. L. (1930). Knowledge and purpose as habit mechanisms. *Psychological Review, 6,* 511-525.

Kennedy, J., and Eberhart, R. C. (1995). Particle swarm optimization. *Proceedings of the International Conference on Evolutionary Computation*, Perth Western Australia. Invited paper.

Kennedy, J (1996). Social optimization of continuous-valued networks. Presentation to the International Sunbelt Conference on Social Networks, Charleston, South Carolina, February, 1996.

Kosko, B. (1992). *Neural networks and fuzzy systems: A dynamical systems approach to machine intelligence.* Englewood Cliffs, NJ: Prentice Hall.

Leavitt, H. J. (1951). Some effects of certain communication patterns on group performance. *Journal of Abnormal and Social Psychology, 46*, 38-50.

Markus, H. and Zajonc, R. B. (1985). The cognitive perspective in social psychology. In G. Lindzey and E. Aronson (Eds.), *Handbook of social psychology, volume 1: Theory and method.* New York: Random House.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning internal representations by error propagation. In Rumelhart, D. E. and J. L. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition. Vol. 2: Foundations.* Cambridge, Massachusetts: The MIT Press.

Thorndike, E. L. (1911). *Animal intelligence: Empirical studies.* New York: MacMillan.