(Extended Abstract)

Hinton & Nowlan's computational Baldwin effect revisit: Are we happy with it?

Brest State Technical University
Moskowskaja 267, Brest 224017 Republic of Belarus
akira@bsty.by

A-needle-in-a-haystack problems

The problem of looking for only one or only a few objects hidden in a huge search space is sometimes called *a-needle-in-a-haystack* problem. It has long attracted and still is attracting many of us in computer science/technology communities. Let's name a few.

- Crammer & Chechik (2004) proposed such a technique defining the problem as *"The problem of finding a small and coherent subset of points in a given data which sometimes referred to as one-class or set covering,..."*
- Joshi et al. (2001) pointed out yet another but a similar situation, writing, *"The traditional evaluation metric of accuracy is not adequate when the target-class is rare. If the class is very rare, say 0.5 %, then predicting everything to be of non-target-class can also achieve very high accuracy level of 99.5%."*

  The technique proposed exploits two *phases* called *positive* & *negative*. Hence they call it *PN-rule*. Weiss (2004) summarized this method as, *"This approach identifies regions likely to contain needles in the first phase and then learns to discard the strands of hay within these regions in the second phase."*
- Weiss (2004) also discusses, *"The role that rare classes and rare cases play in data mining,"* citing an interesting example of a *"machine learning technique to detect oil spills from satellite images (Kubat et al. 1998)."*
- Sabhnani et al. (2003) applied nine different machine learning techniques to the KDD-cup-1999 dataset (Stolfo et al. 1999) to know how these techniques detect network intrusions. In the dataset, data for four categories of intrusions are given, together with data for normal transactions. It was shown that the two out of four categories were all immune to all the nine methods while the other two are easy to be detected. We hypothesized that this is due to these two categories of attacks are like a needle in a haystack of a normal and other categories of attacks (Imada, 2006).
- Among others, the most popular issue on this topic these days is probably a searching for needles in a huge hay of world-wide-web resources, and designing such search-engines. See, for example, (Makris et al. 2006).
- In software engineering community, this issue is also explored. See, for example, (Whitaker et al. 2004).

Hinnton & Nowlan's model – Computational Baldwin Effect

In their seminal paper, Hinton & Nowlan (1987) showed us an elegant experiment which they called *" an evolution with the Baldwin effect in computers."* The purpose was to show a new method to search for a unique object located in a huge search space, and thus called *a-needle-in-a-haystack* problem. The method is by evolving a population of candidates of the solution like a usual evolutionary search, but what was proposed was to exploiting individual's *lifetime-learning* in addition to the standard operation for reproduction. Hence, the proposal is said to be the first suggestion of a computational Baldwin Effect. Since then we have had a fair amount of application proposals of this technique. Even these days. See, for example, (Mills & Watson, 2006). The topic still includes open issues and some of them are explored in this paper.

Hinton & Nowlan's model is quite simple. That is, with a needle being a unique configuration of 20-bit binary string, all other 20-bit configurations constitute a haystack. If we'd just imagine a 20 bit binary PIN code instead of usual 4 digit decimal one, it would be easy to understand what is the needle here.

Note that if we apply a standard genetic algorithm starting with a population of chromosomes with their genes being either 0 or 1 at random, then the fitness of each individual is always zero unless the individual is ultra lucky to be coincidentally identical to the needle. So the fitness landscape is everywhere flat island of altitude zero except for the only one point. Impossible to evolve.

Following Hinton & Nowlan's experiment and a more clear specification of the experimental condition suggested by Mills & Watson (2005), we examined the experiment. We create a population of 1024 chromosomes of 20-bit whose genes are either 0, 1, or 9. Genes are determined at random with a probability of 0.25 for 0's and 1's and 0.5 for 9's. A needle here is the all 1 chromosome, without a loss of generality. Each chromosome is given a chance of 1000 trials each time with its 9 being randomly replaced either with 0 or 1, and check if it matches the needle or not, which is called *lifetime-learning*. If it matches the needle at n-th trial, the fitness is given as $(1000 - n)$. Then we evolve this population from one generation to the next by *fitness-proportionate-selection*, *one-point-crossover* and *mutation* with the probability of 1/20.

Can we observe such an elegant result? See Fig. 1 in which two results: (i) simple random search and (ii) search with lifetime learning are shown. Although in both cases, the number of steps until one of the individuals finds the needle is $O(N)$, what a tremendously dramatic enhancement in efficiency it looks like!

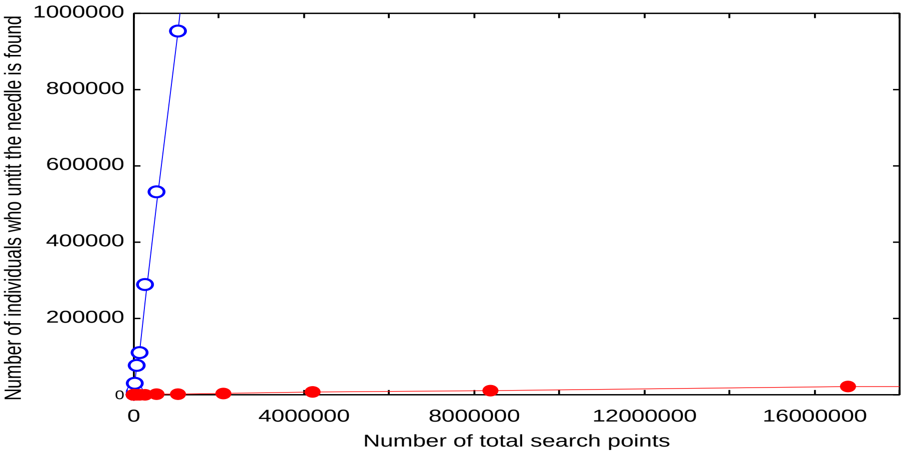

Figure 1: Number of individuals necessary for the needle to be reached for the first time vs. the number of total search points. A result of random walks (blank circles) and search by 1000 learnings during individual's lifetime (filled circles). The data are averaged after 100 runs. Complexity is both linear, but what a dramatic difference!

But, alas, a reality is that the number of evaluations of whether phenotype matches needle or not is similar in both cases. Anyway, by exploiting this method, it is still realistically impossible to reach the needle of, say, 30-bit binary string even with a currently available fastest personal computer.

Watson & Pollack (1999) wrote in their paper *""Hinton and Nowlan choose the population size, number of lifetime trials, and number of variables in the problem carefully so as to make it most unlikely that genetic variation alone will find the solution but very likely that lifetime variation will,"* our observation, however, was not so optimistic, as we will see later.[1]

---

[1] This is omitted in this extended abstract.

Why should we continue when the lifetime learning already find the needle? This is what we wondered when we read the original experiment. A possible answer was given by Mills & Watson (2005) in which they argued: *"This model is not intended to show any engineering advantage* [*but a biological viewpoint.*]*"* Then they went on, *"To remove the assumption of learning phenotypes, we evaluate fitness as the mean fitness of the lifetime phenotypes, rather than the number of trials remaining after the phenotypic solution is first found. This means that the organisms do not have to recognize their own success (as is implicit in Hinton & Nolan's model.)"* [2]

We are wondering how their *"taking an average of all zero fitness indiviuals,"* is meaningful, but anyway, why don't we expect more to other biological evolution model such as *Lamarckian inheritance* rather than sticking to the Baldwin Effect.

Is the Lamarckian inheritance computationally plausible? Turney (1996) wrote, *"Lamarckism requires an inverse mapping from phenotype and environment to genotype. This inverse mapping is biologically implausible."* And further assumed by describing, *"Perhaps Lamarckian evolution is superior to the Baldwin effect, when we are attempting to solve problems by evolutionary computation."*

Then, why not trying this in our Hinton & Nowlan's model. Though Turney strangely wrote the above statement once more as one of the conclusions, *"Lamarckian evolution requires an inverse mapping from phenotype and environment to genotype ... We believe that computing this mapping is intractable in general,"* we cannot agree this, at least in our Hinton & Nowlan's model. All we need is remap some of the successful 0's and 1's in the phenotype to the corresponding genes of 9 in the genotype with a certain probability, though we don't know if this is computationally or biologically meaningful or not.

Can a learning enhance a Robot random navigation?

What about a different scenario? We now take a rook at a problem of robot navigation in a two-dimensional grid. The task of the robot is to look for an object hidden somewhere in the grid. The robot has no idea of where. This is a two-dimensional version of a-needle-in-a-haystack problem. With random walk, the robot can eventually find the needle after an average of $O(N^2)$ steps, assuming $N \times N$ grid. Then question is, "A learning scheme can reduce the number of steps via a series of trial-and-errors afterwards?"

Quantum computation can search a needle with more efficiently?

Since Grover's (1997) assertion that quantum mechanics helps in searching for a needle in a haystack with $O(\sqrt{N})$ steps while classical computer requires $O(N)$ steps,[3] lots of approaches exploiting a *quantum random search* have been proposed. See, for example, (Shenvi et al. 2003). As for searching a space by a mobile robot, Beninof (2002) proposed a *quantum robot.* It might be interesting to see what Beninof wrote:

> *"For this initial memory state all 2^N searches are carried out coherently. Since the path lengths range from 0 to 2^N, the quantum robot can search all sites of R and return to the origin in $O(N \log N)$ steps. Since this is less than the number of steps, $O(N^2 \log N)$, required by a classical robot, the question arises if Grover's algorithm can be used to process the final memory state to determine the location of S. If this is possible, the overall search and processing should require $O(N \log N)$ steps which is less than that required by a classical robot."*

At the same time, however, we must notice that search for a point in d-dimensional hypercube by a quantum random walk is $O(\sqrt{n})$ for $d \geq 3$ and $O(\sqrt{n \log^5 n})$ for $d = 2$ (Aaronson et al. 2003). It would be more effective only in a dimension higher than 3.

---

[2] The key to understand this issue is while genotype in the Hinton & Nowlan's model includes flexible genes 9's – such as (10090011001909100199), its phenotypes are the ones whose 9's are all replaced with either 0 or 1 for a lifetime learning – like (10010011001100100111). Hence one genotype has a maximum of 1000 different phenotypes.

[3] As N is the number of points in search space in his equation, it is $O(N^2)$ in our context.

Concluding remarks

We have shown a kind of survey of the topic on *a-needle-in-a-haystack problem.* In a meanwhile, we had an interesting discussion between two papers Yu & Miller's (2002) *"Finding needles in haystacks is not hard with neutrality."* vs. Collins' (2005) *"Finding needles in haystacks is harder with neutrality."* As we mentioned at the very beginning of this paper, we have lots of proposals of innovative techniques to find a *needle* efficiently. What we have to be careful about, however, will be to avoid an effect of *like-to-hear-what-we-would-like-to-hear.* Most techniques are still vulnerable when they are applied to a more scaled-up circumstance than when it was designed. A success in a small scale experiment is not necessarily a *royal road* to a success in a large scale situation in the real world. This is what we wanted to emphasize in this paper, hoping this article would be a good *prelude* to the researches on this issue.

Acknowledgment

This topic was motivated by a comment from the floor when I talked in a seminar held in Vilnius in 2003 about evolutionary computations in general which included the topic of the Baldwin Effect. The comment insisted, *"Such a search will never be plausible,"* mentioning the Baldwin Effect. Though still I could say, "But it smoothes the fitness landscape, anyway," what he pointed out was correct in a sence written in this article. I now thank you for his comment.

Reference

Aaronson, S., and A. Ambainis (2003) "Quantum Search of Spatial Regions." In Proceedings of IEEE Symposium on Foundations of Computer Science, pp. 200–209.

Benioff, P. (2002) "Space Searches with a Quantum Robot." Contemporary Mathematics, Vol. 305, pp. 1–12.

Baldwin, J. M. (1896) "A New Factor in Evolution." American Naturalist, Vol. 30, pp. 441-457, 536-554.

Collins, M. (2005) "Finding Needles in Haystacks is Harder with Neutrality." In Proceedings of Genetic and Evolutionary Computation Conference, pp. 1613–1618.

Crammer, K., and G. Chechik (2004) "A Needle in a Haystack: Local One-Class Optimization." In Proceedings of International Conference on Machine Learning. http://citeseer.ist.psu.edu/crammer04needle.html

Grover, L. (1997) "Quantum Mechanics Helps in Searching for a Needle in a Haystack." Physical Review Letter, Vol. 79, pp. 325–328.

Hinton, G. E., and S. J. Nowlan (1987) "How Learning can Guide Evolution: Adaptive Individuals in Evolving Populations." Models and Algorithms, Addison-Wesley Longman, pp. 447–454.

Imada, A. (2006) "How Many Parachutists will be Needed to Find a Needle in a Pastoral?" In Proceedings of the International Conference on Neural Networks and Artificial Intelligence, pp.53–60.

Joshi, M. V., R. C. Agarwal, and V. Kumar (2001) "Mining Needles in a Haystack: Classifying Rare Classes via Two-Phase Rule Induction." ACM SIGMOD (Special Interest Group on Management of Data) Record, Vol. 30 (2), pp. 91–102.

Kubat, M., R. C. Holte, and S. Matwin (1998) "Machine Learning for the Detection of Oil Spills in Satellite Radar Images." Journal of Machine Learning Vol. 30, pp. 195–215.

Makris, C., Y. Panagis, E. Sakkopoulos, and A. K. Tsakalidis (2006) "Efficient and Adaptive Discovery Techniques of Web Services Handling Large Data Sets." Journal of Systems and Software, Vol. 79(4), pp. 480–495.

Mills, R, and Watson R. A. (2005) "Genetic Assimilation and Canalization in the Baldwin Effect." In Proceedings of the European Conference on Artificial Intelligence, pp. 353-362.

Mills, R., and R. A. Watson (2006) "On Crossing Fitness Valleys with the Baldwin Effect." In Proceedings of Artificial Life X, pp. 493–499.

Sabhnani, M., and G. Serpen (2003) "Application of Machine Learning Algorithms to KDD Intrusion Detection Dataset within Misuse Detection Context." In Proceedings of the International Conference on Machine Learning: Models, Technologies and Applications, pp. 209–215.

Shenvi, N., J. Kempe, and B. Whaley (2003) "A Quantum Random Walk Search Algorithm." Physical Review A, Vol. 67, pp. 052307–052318.

Stolfo, S. J., F. Wei, W. Lee, A. Prodromidis, and P. K. Chan (1999) "KDD Cup Knowledge Discovery and Data Mining Competition." http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

Turney, P. (1996) “Myths and Legends of the Baldwin Effect” In Proceedings of International Conference on Machine Learning, pp. 135–142.

Watson, R. A., and J. B. Pollack (1999) ”How Symbiosis Can Guide Evolution.” In Proceedings of European Conference on Artificial Life (Advances in Artificial Life), pp. 29–38.

Weiss, G. M. (2004) “Mining with Rarity: A Unifying Framework.” ACM SIGKDD (Special Interest Group on Knowledge Discovery and Data Mining) Explorations, Vol. 6(1), pp. 7–19.

Whitaker, A., R. S. Cox, and S. D. Gribble (2004) ”Configuration Debugging as Search: Finding the Needle in the Haystack.” Journal of Operating Systems Design and Implementation 2004, pp. 77–90.

Yu, T., and J. Miller (2002) “Finding Needles in Haystacks is Not Hard with Neutrality.” In Proceedings of EuroGP 2002, Lecture Notes in Computer Science Vol. 2278, Springer, pp. 13–25.