Croatian Journal of Philosophy
Vol. XVIII, No. 54, 2018

How Gruesome are the No-free-lunch Theorems for Machine Learning?

DAVOR LAUC
University of Zagreb, Zagreb, Croatia

No-free-lunch theorems are important theoretical result in the fields of machine learning and artificial intelligence. Researchers in this fields often claim that the theorems are based on Hume's argument about induction and represent a formalisation of the argument. This paper argues that this is erroneous but that the theorems correspond to and formalise Goodman's new riddle of induction. To demonstrate the correspondence among the theorems and Goodman's argument, a formalisation of the latter in the spirit of the former is sketched.

Keywords: Induction, the problem of; No-free-lunch theorems, The new riddle of induction.

1. *Introduction*

Right from its beginning, the development of artificial intelligence and machine learning prompted many interesting philosophical debates and provoked some interesting philosophical questions. On the flip side, researchers in these fields often encounter or rediscover classical philosophical problems. One such case is the identification of the no-free-lunch (NFL) theorems—the famous negative results in machine learning—as a reiteration or even a formalisation of the Hume's problem of induction. The distinguished machine-learning researchers like Christophe Giraud-Carrier and Pedro Domingos state, respectively:

> It then becomes apparent that the NFL theorem in essence simply restates Hume's famous conclusion about induction having no rational basis... (Giraud-Carrier 2005)

and:

> ...This observation was first made (in somewhat different form) by the philosopher David Hume over 200 years ago, but even today many mistakes in machine learning stem from failing to appreciate it. (Domingos 2012)

Even the originator of the first form of the NFL theorems, David Wolpert, fifteen years after he proved the theorem, joins the information cascade and claims that:

> …these original theorems can be viewed as a formalisation and elaboration of concerns about the legitimacy of inductive inference, concerns that date back to David Hume… (Wolpert 2013)

This paper argues that the NFL theorems, although vaguely connected to the classical philosophical problem of induction, do not restate the Hume's problem, but rather the associated Nelson Goodman's argument.[1] We claim that the NFL theorems are closely related to Goodman's new riddle of induction (NRI), to the extent that they are one possible formalisation of the riddle. Additionally, we would like to pose the question of the relevance of the NFL to the vast philosophical discussion on NRI, as the relationship is yet to be researched. The related, reversed, the issue is the relevance of NRI to NFL and the question as to whether the machine-learning community could benefit from the almost 70 years of fruitful discussion about Goodman's argument.

1.1 No-Free-Lunch Theorems

The first form of NFL theorem was proven by Wolpert and Macready, 1992, in the context of computational complexity and optimisation research (Wolpert and Macready 1995; Wolpert 1992). He later proved the variant of the theorem for the supervised machine learning (Wolpert 1996). For the sake of our argument, we will sketch the proof of the simplified version of the theorem for supervised learning based on the work of Cullen Schaffer (Schaffer 1994).

In the simplest, discrete settings of the machine learning of a Boolean function, training data X consists of the set of binary vectors representing a set of attributes that are true or false for each instance of the binary function—concept. Each vector is labelled as a positive or negative example of a concept we want to learn. The machine-learning algorithm L tries to learn a target binary function y; a true concept from this set of examples. Training dataset is always finite with some length n, and the relative frequency of data feed to L is defined by probability distribution D. In a context more familiar to the philosophers, this problem of machine learning can be seen as a guessing a true form of a large n-ary truth function from the partial truth-table, where most of the rows are not visible.

The key performance indicator of a machine learner is a generalisation performance, with the accuracy of the learner found within the data outside the training dataset. Modern machine-learning algorithms can easily "memorise" data from the training dataset, and perform poorly on the "unseen" data, leading to the problem known as overfitting. So,

[1] We suppose that the new riddle is a different issue from the classical problem of induction, what is the received position with a few notable exceptions like (Magnus 2006).

the success of the learner is measured by how well it will generalise, and how well it performs on the novel data. In the simple setting of binary concept learning, the baseline of the generalisation accuracy of a learner, *GP*(*L*) is the random guess, with the accuracy of the novel data being 0.5. This is the performance we will expect on average if we use the toss of a coin to decide, for an unseen example, whether it belongs to our target concept or not. Clearly, we want any learner to perform better than this.

The NFL theorem claims that, for any learner *L*, given any distribution *D* and any *n of X*

$$\frac{\sum_{f \in Y} GP(L)}{|Y|} = 0.5$$

where *Y* is a set of all target functions, all possible concepts that can be learned.

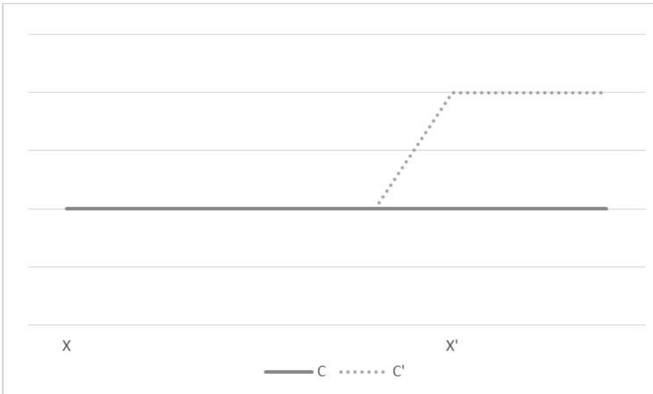
So, the theorem states that, on average, the generalisation performance of any learner is no better than random guessing. All learning methods, from the simple decision trees to the state-of-the-art deep neural networks, will perform equally when all possible concepts are considered.

This result, unanticipated at least on the first sights, does bear some resemblance to the discrepancy between the results of the argument and our expectations in the case of the Hume's argument. However, it does not claim that we cannot learn anything from the training data or experience, but that we can learn everything, which is, arguably, the point of Goodman's argument. The resemblance to the NRI will be more evident from the sketch of the proof of the NFL theorem. The basic idea is very simple: for any concept that the learner gets right, there is a concept that it gets wrong or, in Goodman's lingo, for every "green" concept there is a "grue" concept. The "grue" concept is constructed similar to the NRI argument, in that it agrees on all observed data—data in the training dataset—with the "green" concept, and it is bent on all non-observed data.

More formally, for every concept *C* that *L* learns to classify well—say it classifies *m* novel examples accurately—there is a concept *C′* that *L* learns where all *m* examples will be misclassified. *C′* is constructed as follows:

$$C' = \begin{cases} C \ if \ x \in X \\ \neg C \ if \ x \notin X \end{cases}$$

Visually, this simple construction of the concept *C′* corresponds to the Wittgenstein-Goodman "bent predicate" (Blackburm 1984), where X represents observed data (training dataset) and X' unobserved data.



From the perspective of the main measure of the success of the learning—generalisation accuracy, for every accuracy improvement *a* over the baseline for a concept *C*, there is a concept *C'* that will offset the improvement of the accuracy by –*a*. Consequently, the improvement in accuracy for any learner over all possible concepts is zero. It is possible to generalise this result to the more general learning settings, and many extensions of the theorem are proven (Joyce and Herrmann 2018; Igel and Toussaint 2005).

1.2 *Is the No-free-lunch Theorem the New Riddle of Induction?*

Although NFL bears a strong resemblance to NRI, it seems worthy to analyse differences and similarities between these two results. Let's start with the similarities. Both arguments are about inductive inference, about inferring from the known to unknown, and from the observed to the unobserved data. Both arguments imply that there are too many inductive inferences that can be inferred. Furthermore, both arguments seem to draw empirically inadequate conclusions, contrary to scientific practice and common sense expectation. Nobody is expected to conclude that all emeralds are grue, and neither that random guess is an inductive strategy as good as any other.

The key resemblance is in the construction of the arguments, the split of the evidence and the bend in the unobserved data. In most of the NRI arguments, we split the evidence into observed and unobserved (sometimes to some point in the future). Equally, in the NFL, data are split into observed, training dataset and the unobserved data to which the learning algorithm should generalise. In both arguments, the other counter-concept, *grue* or *C'*, is constructed in the same manner. It agrees on the observed data and bends on the unobserved data.

Regarding the differences, firstly, there is a difference in the argument contexts. NRI was made in the philosophical, theoretical context of the logic of confirmation and pragmatic vindication of induction, while the NFL was made in the technological context of artificial intelligence and computing. The aim of the arguments also differs, at least at first glance. The intention of NRI, at least in Goodman's initial form (Goodman 1946; Goodman 1983), was to recognise one of the problems in the logic of confirmation—the demarcation between projectable and non-projectable predicates. On the other hand, the objective of the NFL was to demonstrate that there is no single best algorithm, initially for the optimisation and search, and later for supervised learning.

The biggest difference seems to be in the scope of quantification. The no-free-lunch theorem quantifies over all learners and all concepts, while Goodman's argument seems to be about constructing one particular example. However, NRI can be reformulated to have a similar quantificational structure as the NFL.

2. *The New Riddle of Induction as the No-free-lunch Theorem*

Goodman's argument, at least in one of its interpretations, can be rephrased in the NFL fashion as follows. Let's define P, the degree of projectability, as the generalisation accuracy in the settings of learning Boolean function. Let's define L to be a language, understood informally as a frame of reference or level of abstraction (Floridi 2008). It is also possible to define language using a more formal framework like a web ontology (McGuinness 2004) or the formal concept analysis (Ganter 2012). Finally, let's define I as an inductive strategy or, as in Goodman's original argument the logic of confirmation. Then, we can state the new riddle as:

$$\forall I \; \forall L \frac{\sum_{x \in L} P(x)}{|L|} = 0.5$$

Claiming that, for any inductive strategy, the degree of projectability over all possible languages is 0.5, what is zero improvement over a random guess. The formal condition, by the NFL condition, is that the languages are unrestricted or that they are closed under permutation (Schumacher 2001). The proof of such stated NRI would be the same as the proof for NFL—for every concept C with a degree of projectability p, there is a concept C' with the degree of $-p$, and for every "green" concept there is a "grue" concept.

3. *Discussion and Future Work*

The takeaway of this formalisation would be one of the lessons that Goodman has taught us—the importance of the language for the induction, or the impossibility of empirical investigation without some predefined language that we bring to the process. It is interesting to compare this with the conclusion that the same researcher from the machine-learning community draws from the NFL—there is no learning without bias, there is no learning without knowledge (Domingos 2015).

If we accept the above formalisation as one of the possible interpretations of the NRI argument, it would be interesting to examine how the NFL can contribute to the NRI exploration. One of the approaches is to investigate how different NFL conditions apply to NRI. For example, if we restrict a set of the concepts to some subset of all those possible, it has to be closed under permutation (C.U.P.) for NRI to hold (Schumacher 2001), and the fraction of such subsets is tiny. Another potential route of exploration is to research the relevance of non-uniform distribution of target functions/concepts for arguing that NRI works (Igel and Toussaint 2005) and results that the NRI does not extend to the continuous domains (Auger 2007).

On the flip side, it would also be interesting to explore possible "solutions" to the NFL theorem from the NRI perspective. Is it possible to use Goodman's pragmatic solution in limiting the NFL by formalising his concept of entrenchment? It would also be interesting to research additional constraints on NFL languages using Davidson's approach to NRI (Davidson 1966). Finally, one of the biggest social and ethical problems of machine learning, especially deep learning, is the problem of the interpretability of models. It would be interesting to research whether the philosophical explorations in the NFL could help in this direction.

References

Auger, A. a. 2007. "Continuous lunches are free!" *Proceedings of the 9th annual conference on Genetic and evolutionary computation.* ACM.

Davidson, D. 1966. "Emeroses by other names." *The Journal of Philosophy* 63 (24): 778–780.

Domingos, P. 2012. "A few useful things to know about machine learning." *Communications of the ACM* 55 (10): 78–87.

Domingos, P. 2015. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World.* Hachette UK.

Floridi, L. 2008. "The method of levels of abstraction." *Minds and machines* 18 (3): 303–329.

Ganter, B. a. 2012. *Formal concept analysis: mathematical foundations.* Springer Science.

Giraud-Carrier, C. a. 2005. "Toward a justification of meta-learning: Is the no free lunch theorem a show-stopper." *Proceedings of the ICML-2005 Workshop on Meta-learning.*

Goodman, N. 1946. "A Query on Confirmation." *The Journal of Philosophy* 43 (14): 383–385.

Goodman, N. 1983. *Fact, fiction, and forecast.* Cambridge: Harvard University Press.

Igel, C. and Toussaint, M. 2005. "A no-free-lunch theorem for non-uniform distributions of target functions." *Journal of Mathematical Modelling and Algorithms* 3 (4): 313–322.

Joyce, T. and Herrmann, J. 2018. "A Review of No Free Lunch Theorems, and Their Implications for Metaheuristic Optimisation." *Studies in Computational Intelligence* 744: 27–51.

Magnus, P. D. 2006. "What's new about the New Induction?" *Synthese* 148 (2): 295–301.

McGuinness, D. L. 2004. "OWL web ontology language overview." *W3C recommendation* 10 (10).

Schaffer, C. 1994. "A conservation law for generalization performance." *Machine Learning Proceedings* 1994: 259–265.

Schumacher, C. M. 2001. "The No Free Lunch and Description Length." In E. G.-M. L. Spector. *Genetic and Evolutionary Computation Conference (GECCO 2001).* San Franisco: 565–570.

Wolpert, D. 1992. "Stacked generalization." *Neural networks* 5: 241–259.

Wolpert, D. 1996. "The Lack of A Priori Distinctions between Learning Algorithms." *Neural Computation* 1341–1390.

Wolpert, D. H. 2013. "Ubiquity symposium: Evolutionary computation and the processes of life: What the no free lunch theorems really mean: How to improve search algorithms." *Ubiquity, December 2013.*

Wolpert, D. and Macready, W. 1995. "No Free Lunch Theorems for Search." *Technical Report SFI-TR-95-02-010 (Santa Fe Institute).*