Beauty in Machine Learning: Purpose and Enlightenment

Anonymous Author(s)
Affiliation
Address
email

Abstract

Drawing on empirical research and theories of beauty from psychology, and on philosophical investigations into the role of beauty in scientific research, I argue that aesthetic considerations are valuable to machine research in two modes. First, as a statistical and computational criterion for selecting theories (i.e., models), which leads to quantifiable benefits and appears to be largely automatable. These notions of beauty are largely the same as "objective" notions of beauty that are closely tied to human sensory stimulation. Second, somewhat less tangibly and less automatable, as desiderata for our models, methods, and scientific literature. These are manifestations of more complicated notions of beauty that are modulated by experience accumulated over time.

What is the role of beauty in machine learning (ML) research? Much of what has been written about the role of aesthetics in scientific inquiry considers its role in *theoretical* science. In contrast to a theoretical physicist, who aims to describe some physical phenomenon, a ML theoretician aims to describe the effectiveness of a human-designed computational method for analyzing data. Moreover, applied ML researchers are engaged in a type of automated theory search; other ML researchers design methods for automated theory search. Do the same considerations of the value of beauty apply in all of these activities?

Drawing on empirical research and theories of beauty from psychology, and on philosophical investigations into the role of beauty in scientific research, I argue that aesthetic considerations are valuable to ML research in two modes. First, as a statistical and computational criterion for selecting theories (i.e., models), which leads to quantifiable benefits and appears to be largely automatable. These notions of beauty are largely the same as "objective" notions of beauty that are closely tied to human sensory stimulation. Second, somewhat less tangibly and less automatable, as desiderata for our models, methods, and scientific literature. These are manifestations of more complicated notions of beauty that are modulated by experience accumulated over time.

1 Objective beauty and ML objectives

In practice, the *process* of applied ML research has much in common with that of theoretical science. Just as a theoretical scientist searches for a theory to accurately describe observed phenomena, fitting models to data can be viewed as searching for an accurate theory of a process from which the data arose. That is, a single prediction function h is a theory (or, more commonly, hypothesis), and the collection of all theories under consideration, $\mathcal{H}$, is the not-coincidentally-named "hypothesis space." A single prediction function h^* that best explains the observed data is selected and we have our theory. Does beauty play a role here?

Under review at ICBINB Workshop @ NeurIPS 2021. Do not distribute.

1.1 Purposeful beauty

The physicist Steven Weinberg wrote, "The physicist's sense of beauty is also supposed to serve a purpose—it is supposed to help the physicist select ideas that help us to explain nature" [1]. In other words, beauty serves a theoretical physicist as a selection criterion for models that describe some truth about nature. Do selection criteria in ML have a similarly purposeful beauty?

Thy psychologists Reber, Schwarz, and Winkielman [2] list three features that, based on a large body of empirical evidence, they claim can be considered as "objectively" beautiful features: simplicity/amount of information, symmetry, and contrast/clarity. The purposeful beauty of these features (the first two in particular) can be found throughout ML.

Simplicity. In ML research, the standard "truth assessment" is out-of-training-sample generalization error.[1] Upon closer examination, this has a notion of beauty-in-simplicity built in. The most widely used bounds on generalization error (e.g., PAC-Bayes [3, 4] or bound based on the VC dimension or Rademacher complexity [5]) have the general form:

$$\text{Generalization error} \leq \text{Empirical error} + \text{Complexity} \ . \tag{1}$$

The practical guidance provided by this bound is clear: in order to select the "truest" theory from those that achieve the same empirical error, we should select the simplest. Aesthetic considerations that value simplicity (after empirical performance) might lead to the same selection. To the extent that small generalization error correlates with truth, and simplicity is beautiful, truth is indicated by a purposeful beauty in (1).

This is the wisdom of the adage, "The best explanation is the simplest one," and is the conceptual basis of Occam's Razor, information criteria for statistical model selection (e.g., AIC and BIC) [6], the minimum description-length principle [7], and the maximum entropy principle [8], among others.

The computational benefits of simplicity are numerous. For example, a simpler model typically requires less computation and data to fit and evaluate, and it requires less memory. This can might be quantified by a reduction in relevant quantities: the number of parameters, the spectral norm of a prediction function, the dimension of the relevant data space, the number of bits required to encode the model, the number of observations required to achieve a certain level of performance (e.g., parametric versus non-parametric convergence rates).

Symmetry. One reliable way to obtain a simple predictor is by assessing the data generating process for symmetries that leave the problem invariant. Those symmetries can be built into the model family $\mathcal{H}$ (e.g., CNNs) and/or the training algorithm (e.g., data augmentation). If the assumption is accurate—that is, if the underlying process really does contain those symmetries—then something useful happens: compared to a non-symmetric method, a symmetric method is simpler (lower complexity) and the empirical risk is at least as good. The explanation for this is simple: symmetry decomposes the problem into a part that doesn't matter and can be discarded (the stuff that varies when a symmetry transformation is applied), and a part that does matter (the invariant stuff). The greater the symmetry the more that can be discarded.

The simplification property of symmetry can be quantified in a number of ways: statistical benefits are quantified by generalization bounds specializing (1); in information theoretic terms, invariance is compressible [9]; and CNNs have a smaller number of parameters than their fully connected counterparts [10, 11].

1.2 Objectifying beauty and automating scientists

Perhaps a better name for these aspects of "objective" beauty is "objective function" beauty. To the extent that an aesthetic value can be quantified, it can be incorporated into ML methods as a term in an objective function. Indeed, regularization towards lower complexity has been standard practice for some time, and regularization towards invariance through data augmentation has become standard practice in deep learning in recent years.

These developments beg the question: Can the inclusion of beauty in models for data be automated? Although there are still some problems to solve, progress has been made towards automating the

[1] Truth in this context is understood to mean something more narrowly defined than, say, what is meant when a physical theory is said to be true; more on this below.

inclusion of these "objective function" aesthetic features into ML methods, particularly simplicity and symmetry. The main open problem in the latter case is detecting symmetries from data. Although challenging, it seems a matter of time before sufficiently general and reliable methods are developed; early versions show promise [e.g., 12, 13].

Conversely, is a necessary condition for good automated data analysis—or, more ambitiously, general artificial intelligence—the presence of beauty? The statistical and computational usefulness of "objective function" beauty suggest an answer in the affirmative, not as an end unto itself, but as an indicator of a system well suited to its environment.

As a statistician who works largely on problems with symmetry, should I feel doomed to replacement by an automated counterpart?

2 Experiential beauty

Of course, things are more complicated than objective function beauty. Empirical research in psychology indicates that history (e.g., repeated exposure), prototypicality, learning, deviation from expectations (surprise!), and attribution all play important roles in human perception of beauty [2]. For lack of a better term, I will categorize these as "experiential". Whereas the usefulness of objective function beauty was formalized by statistical and computational quantities, experiential beauty is more complicated; to my knowledge, no comprehensive formalism exists. If one does, it does not appear to be widespread in ML research (though see section 2.3). At present, experiential beauty has a bigger role to play in the aspects of ML research that are still human-powered and not easily quantifiable: model and method design, and theoretical analysis. For these reasons, my analysis here is far more speculative and philosophical.

2.1 Beautiful model design

Even when an aesthetic quality can be quantified and regularized for, humans are often still able to do better. The recent flurry of literature on various species of convolutional neural networks is a good example [e.g., 14, 15, 16]. Data augmentation may be general-purpose, but it can't match the data efficiency and or generalization performance of a neural network that is designed to be symmetry-preserving. This kind of beautiful model design has a long history in probability and statistics, often going under the guise of "inductive bias" [see, e.g., the discussion in 17]. Researchers' ability to design models in this way typically requires mathematical expertise that is not easily automated, though in some cases even that has been sufficiently abstracted so as to yield the ingredients for automation [18].

2.2 Beauty, enlightenment, and simplicity

The mathematician Gian-Carlo Rota wrote, "Mathematicians may say that a theorem is beautiful when they really mean to say that the theorem is enlightening," used as a device "to avoid facing up to the messy phenomenon of enlightenment," or understanding [19, pp. 181-2]. In ML, as in mathematics, we may use "beautiful" to describe a method or a theoretical result that "gives away the secret ... [that] leads us to perceive the actual" [19, p. 182].

In the face of increasingly complex methods driven primarily by performance on benchmark data sets, calls for researchers to gain understanding—and to enlighten others—have become more prominent. A well-known recent example was the 2017 NeurIPS Test of Time Award acceptance speech by Rahimi and Recht [20, 21], which articulated a need for better understanding of the methods researchers develop and use. Rudin [22] has argued a step further: that our methods should be interpretable from the start.

In this sense, beauty as something that gives enlightenment or encodes understanding can play a key part. But enlightenment is messy, partially subjective—how can we achieve it? Reliably, simplicity appears: simple models are more likely to be mathematically and/or computationally tractable, free of complexities that obscure the key to understanding. Perhaps there are other reliable ways to achieve understanding; I don't know of any. This is an important problem.

2.3 Automated scientists, revisited

The experiential aspects of beauty suggest a dynamic learning theory that incorporates objective function beauty with the dynamics of learning about one's environment. Reinforcement learning models starting with [23], in which the reward function is based on changes in an agent's ability to compress its encoding of an object, and more recently with [24], where reward is a change in perceptual fluency, represent efforts towards computational models of beauty. That work has a certain appeal. Is automating the role of humans in ML research just a matter of getting the reward right? Of modeling the experiential aspects of beauty?

Diversity of experience. The discussion of enlightenment in section 2.2 ignores the fact that the degree of understanding obtained by one person may be completely different from the degree of understanding obtained by another. Rota [19, p. 173] gives as an example of beauty "the Galois theory of equations," which in all likelihood I would find incomprehensible. Such variation appears in all aspects of beauty, driven by the variation in individual experience, cultural factors, and so on [2]. Even simplicity and interpretability don't have universally agreed-upon definitions in ML [25].

Supposing that automating the role of humans in ML research is just a matter of getting the RL problem set up correctly, there's still the problem of training enough agents with sufficient diversity of experience. T. S. Kuhn speculated that "individual variability in the application of shared values may serve essential functions in science" [26]. Diversity of background, experience, thought: these are all good things that won't be automated anytime soon. I may be replaced, but the entirety of the research community won't be.

Freedom of pursuit. The philosopher G. J. Morgan cast the choice of which scientific theory to pursue in decision-theoretic terms, as depending on an individual scientist's values (as captured by a utility function) [27], which may or may not include aesthetic preferences. Morgan concluded that ultimately, "[t]here must also be what might be called the spontaneity of individual values" [27, p. 14]: not only that scientists might vary in their aesthetic preferences and the relative value they assign to them, but also that a scientist might not value beauty at all. The freedom to choose one's utility seems uniquely human, unautomatable.

3 Conclusion

Researchers in psychology have studied human perceptions of beauty, and separately, philosophers of science have analyzed the role of aesthetic considerations in science. Both lines of work are relevant to ML research. I have briefly sketched their relevance, and how they might contribute to further developments.

References

[1] S. Weinberg. *Dreams of a Final Theory*. New York: Pantheon Books, 1992.

[2] R. Reber, N. Schwarz, and P. Winkielman. "Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience?" *Personality and Social Psychology Review* 8.4 (2004), pp. 364–382.

[3] D. A. McAllester. "Some PAC-Bayesian Theorems". *Machine Learning* 37.3 (1999), pp. 355–363.

[4] B. Guedj. "A Primer on PAC-Bayesian Learning" (2019). arXiv: `1901.05353 [stat.ML]`.

[5] S. Shalev-Shwartz and S. Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014.

[6] K. P. Burnham and D. R. Anderson. *Model Selection and Multimodel Inference*. Springer-Verlag, 2002.

[7] P. D. Grünwald. *The Minimum Description Length Principle*. MIT Press, 2007.

[8] E. T. Jaynes. "Information Theory and Statistical Mechanics". *Phys. Rev.* 106 (4 May 1957), pp. 620–630. DOI: `10.1103/PhysRev.106.620`. URL: `https://link.aps.org/doi/10.1103/PhysRev.106.620`.

[9] Y. Dubois et al. "Lossy Compression for Lossless Prediction" (2021). arXiv: `2106.10800 [cs.LG]`.

[10] J. Shawe-Taylor. "Building symmetries into feedforward networks". *1989 First IEE International Conference on Artificial Neural Networks, (Conf. Publ. No. 313)*. Oct. 1989, pp. 158–162.

[11] S. Ravanbakhsh, J. Schneider, and B. Póczos. "Equivariance Through Parameter-Sharing". *International Conference on Machine Learning (ICML)*. Ed. by D. Precup and Y. W. Teh. 2017, pp. 2892–2901.

[12] M. van der Wilk et al. "Learning Invariances using the Marginal Likelihood". *Advances in Neural Information Processing Systems*. Vol. 31. 2018.

[13] A. Zhou, T. Knowles, and C. Finn. "Meta-learning Symmetries by Reparameterization". *International Conference on Learning Representations*. 2021.

[14] T. S. Cohen and M. Welling. "Group Equivariant Convolutional Networks". *International Conference on Machine Learning (ICML)*. Ed. by M. F. Balcan and K. Q. Weinberger. 2016, pp. 2990–2999.

[15] T. S. Cohen, M. Geiger, and M. Weiler. "A General Theory of Equivariant CNNs on Homogeneous Spaces". *Advances in Neural Information Processing Systems 32*. Ed. by H. Wallach et al. 2019, pp. 9145–9156.

[16] R. Kondor and S. Trivedi. "On the Generalization of Equivariance and Convolution in Neural Networks to the Action of Compact Groups". *International Conference on Machine Learning (ICML)*. Ed. by J. Dy and A. Krause. 2018, pp. 2747–2755.

[17] B. Bloem-Reddy and. Teh. "Probabilistic Symmetries and Invariant Neural Networks". *Journal of Machine Learning Research* 21.90 (2020), pp. 1–61. URL: `http://jmlr.org/papers/v21/19-322.html`.

[18] M. Finzi, M. Welling, and A. G. G. Wilson. "A Practical Method for Constructing Equivariant Multilayer Perceptrons for Arbitrary Matrix Groups". *Proceedings of the 38th International Conference on Machine Learning*. Vol. 139. 2021, pp. 3318–3328.

[19] G.-C. Rota. "THE PHENOMENOLOGY OF MATHEMATICAL BEAUTY". *Synthese* 111.2 (1997), pp. 171–182.

[20] A. Rahimi and B. Recht. *Reflections on Random Kitchen Sinks*. URL: `http://www.argmin.net/2017/12/05/kitchen-sinks/`.

[21] A. Rahimi and B. Recht. *An Addendum to Alchemy*. URL: `http://www.argmin.net/2017/12/11/alchemy-addendum/`.

[22] C. Rudin. "Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead". *Nature Machine Intelligence* 1.5 (2019), pp. 206–215.

[23] J. Schmidhuber. "Formal Theory of Creativity, Fun, and Intrinsic Motivation (1990–2010)". *IEEE Transactions on Autonomous Mental Development* 2.3 (2010), pp. 230–247.

[24] A. A. Brielmann and P. Dayan. "A computational model of aesthetic value". *43rd European Conference on Visual Perception (ECVP 2021)*. 2021. URL: `psyarxiv.com/eaqkc`.

[25] Z. C. Lipton. "The Mythos of Model Interpretability: In Machine Learning, the Concept of Interpretability is Both Important and Slippery." *Queue* 16.3 (2018), pp. 31–57.

[26] T. S. Kuhn. *The Structure of Scientific Revolutions*. 2nd. Chicago: University of Chicago Press, 1970.

[27] G. J. Morgan. "The Value of Beauty in Theory Pursuit: Kuhn, Duhem, and Decision Theory". *Open Journal of Philosophy* 3.1 (2013), pp. 9–14.