Literary History After Machine Learning

Ryan Healey, New York University, Digital Theory Lab

The class of machine learning models composed of convolutional networks and transformers has yielded a novel explanation for how acts of abstraction work. This discovery asserts a provocative advance in the relationship between two practices that confront the articulation of *abstraction* and *genre*: digital computation and literary history.

The convolutional neural network is the culmination of a long line of research in computer science that stretches from breakthroughs in model design (the multiple, "deep" layers of the 1979 Neocognitron), backpropagation (algorithms fashioned in the 1980s for retroactive error correction), dimensionality reduction (a technique outlined in 2006 for transforming a large, intractable data set into a condensed précis of features), and the hike in computational scale achieved by graphics processing units since 2010. The convolutional neural network is a system designed to *convolve*, or entwine into matrices, the minute features of an input object (say, an image, a sound, a video) into denser, reduced maps. These maps, or filters, then distribute a sense of weighted significance onto patterns found in an object, which enables the network to identify a nexus of features and objects.

The philosopher Cameron Buckner recently fixed upon the convolutional neural network as an intriguing machine that recasts the concept of *abstraction*: the practice of separating qualities from a thing, seeing those qualities in different things, and generalizing those things under a shared name. The standard philosophical account of abstraction depends on acts of subtraction (removing "excessive" detail until you get to the "core" of a thing) or representation (building detail toward a general idea that accommodates all instances of a certain thing). Philosophers sometimes regard these classical accounts of abstraction as lost causes, inevitably contradicted by outlying exceptions. But for Buckner the unique achievement of the convolutional network is how it locates abstractions by oscillating between exemplars and categories, a concurrent movement he calls "bidirectional transformation." The exemplars transform the categories, and vice versa. Buckner labels this phenomenon *transformational abstraction*.

Transformational abstraction is a machine process that realizes what Locke, the principal thinker of abstraction in the history of philosophy, saw as something "beyond the power of human capacity to frame and retain distinct ideas of all the particular things we meet with: [as] every bird, and beast men saw; every tree and plant, that affected the senses, could not find a place in the most capacious understanding." For humans, Locke writes, "for every particular thing to have a name is impossible"; for convolutional networks, the degree of possibility hangs on the limits of computation, which can juggle considerably more "particular things" and "names." Following Buckner's conjecture, the stakes of abstraction take on unprecedented clout: if what has seemed to be a mysterious, quintessentially "human" talent for abstraction has been reverse-engineered by convolution, a new ability to pinpoint and design abstractions has arrived. In convolution, we can see the groundwork for abstraction machines that will read and manufacture esoteric representational maps for tasks that span everything from oncology to credit scoring to literary theory.

This paper proposes that the discovery of transformational abstraction is a spur to construct a new history of abstraction. The type of abstraction done by matrix convolution is

a product of a long history that began in the European eighteenth century and its obsession with personification, "abstract ideas," and tabular classification. But the recent discoveries found in the engineering of abstraction compel us to shake out how we understand those earlier events. Machine learning, as a technical development after centuries of separate abstraction problems and abstraction-machine prototypes, presses a break in how we describe abstraction. Until now, abstraction has been variously described as an act of "medium independence" that transposes the "concrete" into the "abstract"; a cut-making procedure that enigmatically (or, some say, arbitrarily or violently) decides what details go into a name or idea; a generalization that adds up qualities until they sufficiently capture an idea; or a human talent for understanding a categorical structure that is essential and readymade to the world. But if we have a separate type of abstraction, a mechanized physical power *not* peculiar to a magical human ability, then the shape for a different history of abstraction emerges.

For literary studies, the problem of abstraction leads directly to the problem of genre. Genre presents an active paradox in literary theory because of the mistaken conceptual overlap with theories of form, and Derrida's famous rejection of "the law of genre" as a disciplinarian "formless form" that only "orders the manifold within a nomenclature." But the discovery of transformational abstraction contradicts this image of forceful, random, merely appellative decisions. Convolutional nets call for a revised theory of genre that advances Ralph Cohen's theory of genres as "open categories," literary units that invoke the continuity and discontinuity of forms as they transform in time. For Cohen, one of the difficulties in theorizing genre was "the failure to distinguish changes within a norm and changes of a norm … the [oscillation between the] development and disintegration of the norm." A new understanding of abstraction as the convolution of almost-coincident points solves this problem and overhauls the traditional relationship of similarities and differences that have constituted our accounts of genre. A major consequence of this resolution is that the study of the dynamics of abstraction foreshadows a series of new challenges for literary history to transpose its expertise in the part–whole relations of writing systems into those of other physical systems.

References

Cameron Buckner, "Empiricism Without Magic: Transformational Abstraction in Deep Convolutional Neural Networks," *Synthese* (2018), no. 195, pp. 5339–72

—— "Moderate Empiricism and Machine Learning," in *Deeply Rational Machines: What the History of Philosophy Can Teach Us about the Future of Artificial Intelligence* (Oxford: Oxford University Press, forthcoming 2023)

—— "Understanding Adversarial Examples Requires a Theory of Artefacts for Deep Learning," *Nature Machine Intelligence* 2 (2020), pp. 732–6

Ralph Cohen, "History and Genre," New Literary History, vol. 17, no. 2 (Winter 1986), pp. 203–18

Jacques Derrida, "The Law of Genre," *Critical Inquiry* (Autumn 1980), vol. 7, no. 1, pp. 55–81

Ryan Healey et al., “The Uses of Genre: Is There an 'Adam Smith Question'?" *Representations*, vol. 149 (2020), pp. 73–102

Rachael Scarborough King, “The Scale of Genre,” *New Literary History*, vol. 52, no. 2 (Spring 2021), pp. 261–84

Matthew Kirschenbaum, “Spec Acts: Reading Form in Recurrent Neural Networks,” *ELH*, vol. 88, no. 2 (Summer 2021), pp. 361–86

John Locke, *An Essay Concerning Human Understanding* (Indianapolis: Hackett, 1996 [1689])

Paul North, *Bizarre-Privileged Items in the Universe: The Logic of Likeness* (Brooklyn: Zone, 2021)

Ashish Vaswani et al., “Attention Is All You Need,” December 2017, arXiv:1706.03762v5

Leif Weatherby and Brian Justie, “Indexical AI,” *Critical Inquiry*, vol. 48, no. 2 (Winter 2022), pp. 381–416.