

John R. Smith
IBM T.J. Watson
Research Center

Lessons in Learning

Machine learning has become an indispensable tool for the multimedia community. It is being applied for content analysis, speech recognition, computer vision, multimedia retrieval, and many more problems.^{1,2} Given large amounts of data, computers using machine learning are able to create rich representations and accomplish impressive discrimination tasks. Yet, the way machines learn still differs significantly from how humans learn. Machine learning generally uses statistical and mathematical techniques that do not have a biological basis. Examples include support vector machines (SVMs), Gaussian mixture models (GMMs), neural nets, and others. However, modeling techniques are still evolving and one day may come closer to those used in human learning.

The way computers receive instruction in machine learning is also different in important ways. Typically, everything (data, concepts, learning problems) is presented at once. In doing so, the computer needs to simultaneously create a range of simple and complex representations and learn to solve easy and hard problems. For example, when given training data for thousands of types of animals, computers must learn to discriminate dogs from alligators (basic) as well as understand the difference between Irish wolfhounds and Scottish deerhounds (advanced). We wouldn't teach our children that way, which is why schools are organized into grades, where early grades focus on simple lessons and higher grades build up to more advanced ideas. Likewise, we need to send the computer to school. We must create appropriate lesson plans or more generally develop curriculum-based approaches to multimedia machine learning.

Figure 1 shows an example framework for curriculum-based multimedia machine learning. As illustrated, images are ordered in terms of the complexity of the content, from simple objects to cluttered scenes. The images are introduced in batches of increasing complexity to allow the computer to develop increasingly

sophisticated representations that it builds on sequentially. Similarly, classification problems are ordered in terms of difficulty to allow the computer to acquire basic discrimination capabilities that become the foundation for advanced problems. Although building confidence is not exactly what matters for the computer, deeply layered learning can use these learned representations and discriminators as building blocks for subsequent levels.

The use of curricula for machine learning is motivated by human and animal learning. The idea of shaping is to schedule a progression of training exercises that establish basic concepts early on, which are then built on to acquire more complex concepts. Shaping has its origins in the work of B.F. Skinner, who discovered that the learning of complex skills improves through successive approximations compared with pure trial and error.³ Given the effectiveness of shaping in human and animal learning, it is reasonable to apply it to machine learning.

The concept of shaping appears in the machine learning literature mainly in two general frameworks. One consists of learning language models and grammars. J.L. Elman showed how a connectionist network performs better in learning grammars when forced to start small and undergo developmental changes that resemble the increase in working memory occurring over time in children.⁴ This was achieved by providing gradually more complex sentences in successive learning stages. A similar approach was developed for learning language models using a deep neural network by Yoshua Bengio and his colleagues.⁵ Shaping has also been explored in robot vision for reinforcement learning. One approach is to start the robot in states that are "close" to the desired goal and then progressively introduce more complex situations that are further away.⁶

Increasing availability of data creates more opportunity to improve multimedia analytics capabilities. And advances in computation make

New Editorial Board Member



Gang Hua is an associate professor of computer science at Stevens Institute of Technology. Before joining Stevens, he worked as a full-time researcher at leading industrial research labs including IBM, Nokia, and Microsoft. His research in multimedia and computer vision studies the interconnections and synergies among visual data, semantic and situated context, and users in the cyber and

physical worlds, which can be categorized into three themes: human-centered visual computing, big visual data analytics, and vision-based cyber-physical systems. Hua has an MS in pattern recognition and intelligence system from Xi'an Jiaotong University and a PhD in electrical and computer engineering from Northwestern University. He has published more than 70 peer-reviewed papers in international journals and conferences. To date, he holds nine US patent and has 13 more US patents pending.

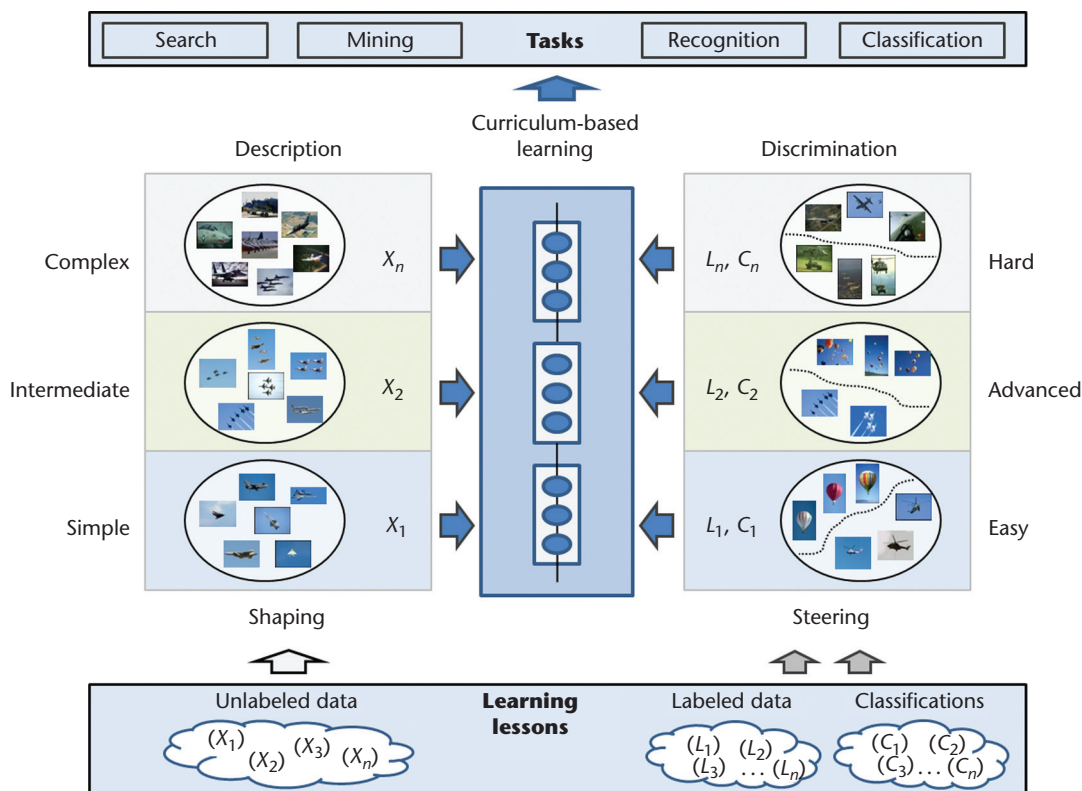


Figure 1. Example framework for curriculum-based multimedia machine learning.

computers capable of metaphorically walking while chewing gum. Nevertheless, computers still need to learn to walk before learning to run. That's why we need to learn how to better structure machine learning through lesson plans and curricula to achieve more effective overall learning. That is lesson number one for us. **MM**

References

1. J.R. Smith, "Minding the Gap," *IEEE MultiMedia*, vol. 19, no. 1, 2012, pp. 2–3.
2. J.R. Smith, "Just the Facets," *IEEE MultiMedia*, vol. 20, no. 1, 2013, pp. 92, 91.
3. G.B. Peterson, "A Day of Great Illumination: B.F. Skinner's Discovery of Shaping," *J. Experimental Analysis of Behavior*, vol. 82, 2004, pp. 317–328.
4. J.L. Elman, "Learning and Development in Neural Networks: The Importance of Starting Small," *Cognition*, vol. 48, 1993, pp. 781–799.
5. Y. Bengio et al., "Curriculum Learning," *Proc. 26th Ann. Int'l Conf. Machine Learning, ACM*, 2009, pp. 41–48.
6. M. Asada et al., "Purposeful Behavior Acquisition for a Real Robot by Vision-Based Reinforcement Learning," *Machine Learning*, vol. 23, nos. 2–3, 1996, pp. 279–303.

John R. Smith is a senior manager of intelligent information management at IBM T.J. Watson Research Center. Contact him at jsmith@us.ibm.com.