The Ever-Growing Importance of AI and Machine Learning in Media and Entertainment

By Steven Bilow

The growth of Over-the-Top (OTT) and Video on Demand (VOD) delivery has increased the volume of media in our industry. Automated captioning, intelligent transcoding, storage and retrieval, automated quality assurance, and cybersecurity needs are rapidly growing. Underlying this are the technologies of machine learning (ML) and artificial intelligence (AI). Recently, a large systems integrator in the industry told a group of professionals that 100% of their clients are asking for ML or other AI technologies as part of new projects.

Automated captioning software is replacing teams of humans. Recommendation engines are part of nearly every service. Our massive volume of rich content can be monetized with increasing flexibility and agility, but only if AI can empower us to manage, manipulate, and store it at scale. We see this as a topic that again is worthy of an entire issue.

Storage capacity and processor performance are now up to the task of executing ML algorithms fast enough for practical tasks like facial expression analysis, virtual character creation, realtime color correction, perceptual modeling, content personalization, and countless other valuable activities.

Our industry is just beginning to explore the vast applications for AI and ML. The articles in this issue cover diverse applications that are no longer theoretical but are now practical. The following articles covering media-oriented applications clearly illustrate this.

In the article *"Realizing Additional Value From Linear Content Using Metadata and Automation,"* Michael Armstrong et al. shows how the BBC uses AI/ML techniques to combine a linear program output and the associated metadata to segment programs into individual stories, which can be searched for, found, and made available on web pages. They demonstrate the use of object-based media (OBM) techniques to allow these stories to be recombined into personalized and potentially interactive experiences.

"Customized Facial Expression Analysis in Video" by Daniel Rüfenacht and Appu Shaji describes a novel method for accurate facial expression recognition in video content. It demonstrates how to successfully classify an extensive selection of custom expressions. The techniques are rooted in the Ekman/Friesen facial action coding system and are thus grounded in robust subject matter expertise. A unique aspect of the solution is that it can remove faces from the dataset that are unsuitable for expression analysis, thus reducing the amount of unusable source data.

In *"Exploring Realtime Conversational Virtual Characters"* by Ha T. Nguyen et. al., we find a technical discussion of what the authors call "the technology components that make up the full cycle of one conversational interaction with a virtual character." This "cycle" progresses from understanding and processing a spoken word to producing a synthesized response that is believable in the context of a character's defined traits. The article details the implementation of a virtual character able to convincingly discuss herself and the Eurovision Song Contest.

In another applied application/practices article *"Understanding Banding—Perceptual Modeling and Machine Learning Approaches for Banding Detection"* by Hojatollah Yegeneh et al., we explore two types of computer-assisted banding detections. The first is called "knowledge-driven" and uses computational models of the human visual system. In this portion of the article, the authors discuss the acquisition, production, distribution, and display of content as well as the interaction between these processes. The second type is referred to as a "data-driven" ML technique that trains a deep neural network using a large-scale dataset.

In 2016, Netflix proposed a technique called "per-title-encoding." Put simply, video content of different complexity levels needs different encoding characteristics to attain a given level of quality. To accomplish this, one must analyze complexity. This is computationally intensive and requires iterative encoding. It, thus, seems like a perfect opportunity to apply ML. In relation to that topic, we present Daniel Silhavy et al.'s article *"Machine Learning for Per-Title Encoding."* Herein, the authors demonstrate how per-title encoding decreases storage and delivery costs while maintaining or improving perceived quality. Supervised multivariate regression algorithms, such as a random forest, are used to predict required quality simplifies complexity analysis compared to standard per-title encoding. This is a rigorous explanation of how and why.

AL and ML are ever-expanding in their practical applications in media and entertainment. We hope that these papers demonstrate that, excite you, and provoke a desire to learn more!

About the Author

Steven Bilow is the senior product marketing manager for video test and synchronization products at Telestream LLC in Beaverton, OR.

Digital Object Identifier 10.5594/JMI.2022.3160259
Date of publication: 6 April 2022

1545-0279/22©2022SMPTE