AI and Machine Learning

By Steven Craig Bilow

Film, television, and streaming delivery of news, sports, and entertainment—along with the associated media asset management, production, and distribution processes—are increasingly information-heavy. Such volume of data makes applications for machine learning (ML) and artificial intelligence (AI) even more important.

The news media overflows with stories as diverse as self-driving cars, video doorbells, recommendation engines, facial recognition, surveillance, and more. But, video, audio, metadata, captioning, and other rich content present extraordinary opportunities to change the way we work. All the issues discussed in other industries, on a daily basis, will affect media businesses as well, arguably more so, because when over-the-top (OTT) and traditional delivery combine, we can access a large portion of the content-consuming audience. That is why SMPTE once again presents an issue on this crucial topic.

ML is based on an algorithmic foundation that is many years old. But the storage, central processing unit (CPU), and graphical processing unit (GPU) performance we have today now provide the necessary computational power to implement real-world solutions to tasks such as facial recognition, video quality assessment, realtime color correction under rapidly changing weather conditions, automated rotoscoping, talent casting based on regional, linguistic, and cultural data, and content personalization. Our industry has long stored massive amounts of data, which is advantageous because it offers countless possibilities.

AI is a massive subject and even ML is only one piece of this larger discipline. Several papers we present here discuss deep learning, which is a further refinement of ML rooted in models based on multilevel neural networks. In a subject area this large, we can make few generalizations. One that we can make, however, is that these are no longer abstract concepts. They are widely deployed. (Just ask Alexa.) The following papers offer powerful media-specific ideas and unique applications. The technology is usable now and these papers demonstrate that.

Here's what to expect.

"Exploring Automated Voice Casting for Content Localization using Deep Learning" by Aansh Malik is a technical paper that explores the use of deep learning to automate what is now a largely manual workflow for casting voiceover talent across languages and cultures. The paper discusses architectures for sequential data processing to create AI-enabled audio-processing workflows. Malik considers ways to leverage developments in deep learning for text-independent speaker verification (TI-SV) to enable computer-aided voice casting.

In the paper "AI News Anchor with Deep Learning-based Speech Synthesis," Kiyoshi Kurihara *et. al.* present a detailed discussion of how they use a deep neural network and speech synthesis to create an artificially intelligent news anchor. Their system is being used to read news automatically and for AI anchors of live programs on NHK. It is in use now and they hope it will serve as a model for other applications.

"Artificial Intelligence: Transforming the Live Sports Landscape" by Amer Saleem and Adrish Bera is a survey of the ways that AI is transforming live sports. For example, identifying specific game constructs, players, events, and actions to generate metadata that leads viewers to the most personally relevant content. This may also be used to automate highlight package creation from both in-game events and viewer preferences. As they say, "AI and ML play a vital role in achieving unprecedented efficiency in sports production, boosting viewership, and increasing ad monetization." This is a fascinating survey of possibilities.

"Towards Designing a Subjective Assessment System for the Quality of Closed Captioning Using Artificial Intelligence" by Somang Nam describes an exciting system for quality assessment of closed captioning. Nam considers both deaf and hard of hearing (D/HoH) audience populations, commonly dissatisfied with current closed-captioning (CC) quality.

Digital Object Identifier 10.5594/JMI.2021.3057688
Date of publication: 6 April 2021

1545-0279/21©2021SMPTE

This paper proposes training an AI to predict the subjective CC quality by learning from human assessment data.

"Automated Brand-Color Accuracy for Real-Time Video" explores a process to ingest live video and to adjust each frame using a machine-learning model trained on a data set consisting of raw and color-corrected videos. The model targets specific values and adjusts only targeted colors. It does so pixel by pixel without shifting surrounding colors. The tool, ColorNet, performs local corrections even as lighting and weather change. It is computationally efficient and outputs a color-corrected feed in realtime. In our brand-centric world, this technology has great potential in any application where color precision is paramount.

We hope you find these papers enjoyable and enlightening!

About the Author

Steven Bilow is the product marketing manager for Tek TV Test and Measurement products at Telestream, Inc. He has more than 30 years experience in the broadcast media industry. Bilow is a senior member of the Association for Computing Machinery (ACM) where he serves on several USACM committees and is a SIGGRAPH Pioneer Member. He is a long-time SMPTE member who first published in the Journal in 2001 and now serves on the Board of Editors. Bilow holds a bachelor's degree in music composition with a minor in film and video from California Institute of the Arts and his graduate work was in computer science with research interests in software metrics at Portland State University.