HUMAN-MACHINE INTERACTION AND INTELLIGENT UNDERSTANDING IN DISSEMINATING VIDEO

ZHANG XIAO*
YANG DELING**

ABSTRACT

In the realm of video, humans use machines to record human histories, beliefs, attitudes, desires, and dreams. Humans and a range of digital media devices are becoming increasingly fused into “human-machines” with trans-humanist lives and existences. Viewers stare at the screen, and the screen of the device “stares” back. Through data collection and analysis, machines can “analyze” and “understand” viewer’s choices of and reactions to videos. Through big data, the internet, and other means, machines are reaching the roots of video creation. Intelligent human-machine interaction in video creation and dissemination is happening. This re-use of data helps people to make decisions about the production and dissemination of video content. Human-machine interaction in video creation and dissemination is inevitable as machines intervene in a networked society and as big data and algorithms intervene in the emotional expressions of video creators.

Keywords: Video, Human-Machines, Simulacra, Agent, Big Data, Artificial Intelligence.

INTRODUCTION

We are accompanied by various digital media devices, and humans and machines are becoming increasingly fused into “human-machines” with trans-humanist lives. The spread of the internet means that video is present in our lives at all times. Viewers stare at the screen, and the screen of the device “stares” back. Smart media devices are constantly collecting viewer data and monitoring viewer behavior. Through the internet and algorithmic analysis, machines can “understand” a viewer’s reaction to a video and generate data. Big data on audiences generated by the internet provides a foundation of data for the creation, dissemination, and production of video. With the rapid development of machine learning and artificial intelligence, machines are relying on the deeper learning of internet video data. This paper will examine the latest developments in big data and algorithms with respect to the creation and dissemination of video, combined with recent case studies in video, artificial intelligence, and deep learning. Through the dual perspectives of audiences and video production companies, this paper

* Associate Professor, Guangzhou Academy of Fine Arts, China. 742717769@qq.com
** Lecturer, Guangzhou Academy of Fine Arts, China. 1229516998@qq.com
will explore how big data and the internet engage with the sources of video culture and how human-machines interact with the future creation and dissemination of video.

**The Big Data Analysis of Video: Viewer Simulacra and Agents in Creation**

Video was invented in the modern humanist context, so video creation and dissemination are centered on the significance of humans. Video is a tool for telling cultural stories and video works allow humans to express their ideas and emotions. Technological advancements related to video have focused on improving the creation, dissemination, and experience of video. As Robert Hughes noted, “People inscribe their histories, beliefs, attitudes, desires, and dreams in the images they make,” and it is people (directors, videographers, and producers) who decide to use machines to create video content.

In humanism, the value of the person is fully respected, and the creation of video has an intensely personal quality. The stories may come from the author or screenwriter’s life experiences and personal thoughts, and the director interprets these stories through his or her own audiovisual language. Through their performances, the actors shape characters with distinctive personalities. The cameraman crafts images with his aesthetic, while the art director creates the visual effects. Through video, we can appreciate personal taste, will, and intellectual freedom. A work with personality draws individual viewers, establishing a pathway for communication, and if the work is appreciated by many viewers, they form an audience that constitutes the foundation of cultural communication. As tools, machines only differ in their functions within video creation. The creation of a video is first and foremost attributed to the creator’s decisions, but a video is also the creative result of an individual’s use of machines based on specialized knowledge.

Interactions with the viewer reflect a kind of respect for the individual’s significance. Those who create and disseminate video attach importance to the viewer’s reaction to the content, and they want to build an interactive, communicative relationship with the audience. Through TV ratings surveys, organizations “obtain continuous information on a sample family or individual, touching on whether they watched TV during a given time slot, and what channels and programs they watched” (Chang, 2014) to provide feedback for video creators. Viewer statistics are analyzed to gauge viewers’ reactions to film and television content, such that film and television production and distribution organizations can tailor their work to the viewing market. The primary goal was to improve viewership and provide other commercial services.

However, with the rise of the digital humanities, big data, algorithms, and intelligent analysis have already intervened in the study of video viewership. “The digital humanities is work at the intersection of digital technology and humanities disciplines” (Drucker, 2013). Through the digital humanities we obtain knowledge using digital technologies; big data, statistics, and algorithms can provide an intelligent analysis of complex cultural and social phenomena. With the spread of the internet to all corners of society, digital media devices continuously convey information to people, and people are faced with the pressure of information overload. Every day, every hour, and every minute, we can choose from a vast amount of video information. YouTube statistics for March 2017 showed that “300 hours of video are uploaded to YouTube every minute” (Fortunelords, 2017) and “almost 5 billion videos are watched on Youtube every single day” (Fortunelords, 2017). At this point, it is impossible for an individual to view all of the video content in the world. Internet videos are viewed differently than traditional
television, as anyone can select and view these videos on a smart phone or tablet anywhere and at any time. Traditional research methods cannot accurately determine individual viewers’ reactions to this kind of video content. Thus, in June 2006, Nielsen began to use cross-media research methods to develop video viewership research for smart phones, tablet computers, digital video recorders (DVRs), internet streaming, and other media devices. In October 2013, Nielsen released its Twitter TV Ratings, which was “the first-ever measure of the total activity and reach of TV-related conversation on Twitter. Nielsen Twitter TV Ratings measure not only ‘authors’—number of people tweeting about TV programs—but also the much larger ‘audience’ of people who actually view those Tweets” (Nielsen,2013).

Big data and algorithms construct an audience database through the collection and categorization of many individual behaviors. These constructed audience databases become data simulacra (Baudrillard, 2006) for individuals and information agents for video creators. Big data and algorithms translate viewer behavior into data and information; these informative and quantifiable methods categorize various kinds of information and obtain data on viewer reactions and behaviors related to video creation and dissemination through intelligent analysis. This material, quantified in data, is the foundation for an institution’s understanding of the individual viewer. A basic understanding of an audience is rooted in the classification of information, so the viewer’s individual value is to provide data for audience information classifications. However, a viewer’s complex personal emotions, which do not fall within the scope of the quantified material of these classifications—and therefore, do not have “tag-able” value as data—are considered noise. As individual feelings and viewer data diverge, human emotions are reduced to data flows from viewers, which concentrate data analysis on constructing an audience; the information obtained from data analysis are the viewers’ feelings. Video creators’ understandings of this data are prioritized over the understanding of the individual: the audience data can represent the “individual.” Thus, online data has replaced the individual, becoming data simulacra for individuals. Video creation based on viewer data simulacra is an interaction between the creator and audience data. The data tells the creator about viewer reactions to video content, providing a market reference for the creation and dissemination of video that is hiding within the experiences, aesthetics, and even intuition of the creators and distributors, becoming an agent of information for the video creators.

Figure 1. Viewer feelings, audience information, and video creation and dissemination decisions
Data simulacra act as “agents” intervening in the creation and dissemination of video, and bring digital technology into humanism; the coexistence of humans and machines is the result of these new core values. At almost all times, people’s lives are closely linked to the rest of the world through various digital devices, creating human-machine lives. People not only receive a large quantity of video data, but they also upload their personal data to the internet; thus, video and personal data become part of the big data flows on the global internet. “Big data is being generated by everything around us at all times. Every digital process and social media exchange produces it. Systems, sensors and mobile devices transmit it” (IBM, 2017). Big data flows are staggering; according to IBM Big Data statistics, the world has generated 2.5 EB of material every day since 2012. (Prior to 2011, the total capacity of the internet was less than 525 EB). In addition, “big data is arriving from multiple sources at an alarming velocity, volume, and variety” (IBM, 2017). The constant arrival of new digital devices has also made data sources more complex. Institutions that create and disseminate video content and rely on traditional methods, such as personal experience and aesthetics, find it very difficult to manage such a large and complex body of information. When faced with market competition pressures and driven by commercial goals, some video companies (such as Netflix) have moved toward a data-driven approach facilitated by algorithms. Data analysis provides an important reference point for creation and distribution decisions, and when combined with the ideas, intuition, and judgment of experts, it is an additional strategy for ensuring a film’s success. The media devices that are with us at all times can provide individual data for algorithms, so these devices play a role in monitoring viewer behavior; the personal data uploaded to the internet constitutes this quantifiable data. In a networked society, machines will become a part of humans and the coexistence of humans and machines replace the individual as the core source of value. Algorithms can also become a part of modern humanism, a new source of meaning hiding within video culture.

Monitoring Viewer Behavior: The Data Simulacra in the Video’s Meaning Network

Human-machines allow the internet to monitor individual behavior, and intelligent analysis can be used to determine an individual’s preferences and habits. Every action that humans take on machines such as smart phones, computers, smart TVs, and online gaming devices leaves a personal trace. All of these traces can be translated into data, which can be collected and collated into databases as a resource. Every viewing choice on a video website, every viewing comment, every online purchase, every piece of information posted on WeChat or microblogs, and every click on the screen leaves data on the internet, including “customer ID, movie ID, rating and the date that the movie was watched” (Marr, 2015). Through the internet, companies can obtain basic information on a user’s hobbies and personal habits. With the spread of streaming media, the internet can dive deeper into viewer behavior data through the smart devices used by viewers, which means the more detailed monitoring of the individual viewer’s behavior. Some video companies are closely analyzing and studying the details of an individual’s viewing behavior. Netflix tracking, for example, is rather detailed, including:

When you pause, rewind, or fast forward; What day you watch content (Netflix has found people watch TV shows during the week and movies during the weekend.); The date you watch; What time you watch content; Where you watch (zip code); What device you use to watch (Do you like to use your tablet for TV shows and your Roku for movies? [...] When you pause and leave content (and if you ever come back); The
ratings given (about 4 million per day); Searches (about 3 million per day); Browsing and scrolling behavior. (Bulygo, 2013)

Netflix is transforming more behavioral details into data, from where and when viewers watch, to choice of content, to viewing frequency and even fast-forward and rewind patterns. "Data such as time of day that movies are watched, time spent selecting movies and how often playback was stopped (either by the user or due to network limitations) all became measurable" (Marr, 2015). Data on small behaviors are the foundation of the algorithmic analysis of individual preferences and habits.

Commercially-driven data collection means that an audience segment constructed by data represents the viewer. Individual reactions to videos are just categorized information within collections of data, and the viewer’s individual value lies in the personal data her or she can provide to the internet. Through algorithms, data such as personal preferences and habits become market information and serve the commercial aims of film production and dissemination. Based on the results of viewer data, companies make market determinations and formulate production and dissemination plans to obtain more views and seek greater commercial benefit (Marr, 2015). Algorithms directly utilize personal data to create commercial value. Audience segments comprised of multiple individual viewers of course have more commercial value, so the collective satisfaction of this audience segment is very important. The value of the individual lies in contributing one’s own data to this audience segment. Although the audience segment does not indicate a specific person (it is simply a temporary collection of individuals who expressed similar preferences for a given film at a given time), recreating the audience’s web data can represent the personal preferences of viewers in this group to meet commercial needs. The unique personal preferences of the viewer are no longer most important; what is important is the ability to input data into the column defined in the company’s data set. Individuals exist in data form within these various data sets. The audience segment and the individual are comprised of data, and commerce and consumerism transform web data into data simulacra for individual viewers.

Machines transform the individual person into a data node on the internet, and numerous nodes, together with machines, constitute the video’s meaning network. Audience segments mean that the viewers are no longer the central concern; the transition from humanism to the digital humanities means that people are no longer focused on themselves, and they are gradually moving toward the fusion of man and machine. The coexistence of humans and machines allows machines to participate in human will. As the limits of human judgment and machine intelligence become vague, the human-centric model is shifting toward the grey area of the fusion of humans and machines. Online communities bring together people from different places around the world to communicate online. Through micro-videos on WeChat and the microblogs, or liking or commenting on video content, humans and machine intelligence influence one another, conveying video information to each other. Viewers’ choices of video can also be influenced by internet data. (e.g., rating or commenting on a film on Rotten Tomatoes can influence a viewer’s choices). Video sites can track viewer browser history, and through a smart analysis of the viewer’s preferences, these sites can automatically recommend videos to the viewer (Marr, 2015). Conversely, viewers’ individual video viewing behaviors can become computer-recorded data that is uploaded to the internet
to influence others. (e.g., comments on Rotten Tomatoes or a video barrage can also influence other viewers). Through the internet, the individual receives constant video information, but the individual is constantly sending personal data out through the internet, such that the individual becomes an information node in the online world. These information nodes are interconnected and trade information with one another; with machines, they intelligently participate in the collective consideration and judgment of the ideas and values in the videos. People, machines, and the global internet together constitute a video’s meaning network.

Data-Driven Video Creation: An Agent in Intelligent Creation and Dissemination in Film Culture

With market competition and the development of artificial intelligence technologies, a data-driven approach can act as an agent in video investment decisions, play a role in video creation, and even directly produce video. The exchange of data between video companies and the audience constitutes a new process of video reproduction: the videos are disseminated in data form on the internet, and by analyzing the audience data for those videos with AI, video companies are provided with possibilities for the next production. Using rational means, algorithms participate in the decision to begin creating a video. Thus, AI and data intervene in the innovations of video creators. With the constant improvements in artificial intelligence’s ability to recognize video content, the machines of the future may even be able to directly create for themselves, independently producing video. Video creators’ ability to formulate plots and characters, as well as their aesthetics, judgments, will, and even creative motivations could be gradually learned and imitated by AI, which will gradually become an agent in the production process. The interactive dissemination of video information between people and between people and machines will develop human-centric video creation and dissemination in the direction of AI creation and dissemination.

Thus, big data and algorithms begin to participate in TV and film investment decisions. Big data and algorithms draw on the massive amounts of data on the internet, especially viewer behavior data, to predict viewership in advance. Based on the probabilities predicted, big data and algorithms can have significant commercial value. Film data and statistics can influence production decisions at film production companies, as they did in Le Vision Pictures’ film Tiny Times (Bangshang You Mingmao, 2014) and Netflix’s web series House of Cards (Bulygo, 2013). Before deciding to invest, Le Vision Pictures took microblogs as a primary data source, and supplemented this data with information from video sites. Data mining can provide a production company with detailed viewer behavior information. The data showed the target audience’s preferences and focuses, and predicted the market value of shooting the script; the data partially acted as an agent in the decision to invest in production and directly influenced the shooting of the film.

In the future, data-driven approaches and artificial intelligence can even serve as direct agents in the creation and production of video. Machine learning technologies for recognizing video mean that machines can actively recognize video content, and video image recognition will develop toward intelligent, automatic recognition. (e.g., at Google: “The Cloud Video Intelligence API (Application Programming Interface) [...] uses powerful deep-learning models, built using frameworks like TensorFlow and applied on large-scale media platforms like YouTube. The API is the first of its kind, enabling developers to easily search and discover video content by providing information about
entities (nouns such as “dog,” “flower” or “human” or verbs such as “run,” “swim” or “fly”) inside video content. [...] These APIs let customers build the next generation of applications that can see, hear and understand unstructured data [...]” (Google Cloud Platform, 2017). The intelligent recognition of video content means that machines can begin to identify the natural, societal, and human content represented in videos, and even gradually understand our “histories, beliefs, attitudes, desires and dreams.” Content recognition is the foundation of video editing, and AI could become a new “brain” for video creation, laying the foundation for intelligent video production in the future.

![Figure 2. The Cloud Video Intelligence API in Google Cloud Machine Learning (Google Cloud Platform, 2017)](image)

Through specially-trained artificial neural networks (ANN), the deep-learning vision system developed by the Computer Science and Artificial Intelligence Laboratory (CSAIL) at MIT has been “trained on YouTube videos and TV shows [...]”, the system can predict whether two individuals will hug, kiss, shake hands or slap five. In a second scenario, it could also anticipate what object is likely to appear in a video five seconds later” (Conner-Simons & Gordon, 2016) Based on the recognition and analysis of internet videos, machines can learn to “understand” the actions and thoughts of a character in a video through algorithms, and even “predict” the next action. The understanding of and familiarity with human behavior is also the foundation of performance in video and its creation. Machines rely on data and internet videos and use algorithms to begin to imitate human “will” in an attempt to determine the character’s behavior.

![Figure 3. The deep-learning vision system can predict what happens next: a handshake (Conner-Simons & Gordon, 2016)](image)
Carl Vondrick, Antonio Torralba, and Hamed Pirsiavash developed a system that generates videos with scene dynamics. By identifying and analyzing the foreground and background of a static image, the system can automatically generate a short video based on that static image. (Vondrick, Pirsiavash & Torralba, 2016) After a static image is entered into the system, the machine relies on web video data to intelligently identify the content of the image, predict movement, and automatically produce a short video based on the predicted movement from that image. Thus, the machine uses web video data analysis and algorithms to create independently. The creation of video could be separated from human will, and video could be created independently by an algorithm; the machine could be an agent with creative “freedom” and begin to produce video on its own.

Figure 4. Conditional video generation (Vondrick, Pirsiavash & Torralba, 2016)

Data-driven video creation shows that the digital humanities can come to more deeply understand nature and society and participate in the development of cultural reproduction. Humanism holds that humans are the source of all meaning, but as people are becoming deeply intertwined with machines, people are changing. As result, humans are no longer the only center. In the context of the digital humanities, the grey area of the human-machine is becoming a new center, creating new sources of meaning. In the context of humanism, video presents people’s observations, experiences, thoughts, and expressions of nature and society, and people rely on themselves to create a source of meaning, using video to create culture. In the context of the digital humanities, video is the reuse of web data filtered through algorithms; it is used to produce new data and constantly expand the amount of information on the internet. Based on their understanding of our algorithms, intelligent machines take meaningless data and produce meaning that suits people. They trade data for data, and the coexistence of humans and machines is the medium of this data exchange. Thus, data-driven video creation is based on the recreation of web video; it is the reproduction of data based on big data and it is a machine’s development of cultural intelligence rooted in video. Data and algorithms have begun to intervene in the sources of video production, which further promotes the development of the digital humanities in future society.
CONCLUSION

Data simulacra and artificial intelligence can serve as agents in video creation. Intelligent dissemination methods mean that video has become an exchange of data over the internet. The fusion of man and machine is the medium for the transfer of this data, which means that the “individual,” whether viewer or creator, becomes a node of information in the global internet. Agents in intelligent video creation and dissemination are layered into human culture through data. By constantly reproducing video that satisfies the market and constantly expands the quantity of big data on the internet, data can become more detailed and algorithms can become more accurate and effective. This in turn will make video productions that satisfy even more people, in an endless cycle. When the creation and dissemination of video by machines intervenes in global meaning networks, algorithms can become sources of meaning and create the cultural stories of video today. People need and even rely on video; because video provides information, entertainment, communication, and knowledge, people are unavoidably influenced by data and algorithms. With the future development of AI, the media devices that come everywhere with us can actively observe, analyze, and study viewers, and machines may even be able to independently consider how to automatically adjust video content to constantly satisfy viewers.

Cultural stories shape societies and people. The cocoon created by video in the digital humanities is developing through the constant augmentation of smart technologies, and with artificial intelligence continuing to learn, it will come to better understand humanity, society, and nature. If the internet creates the future spectacles of society through the increasingly intelligent and deep control of video creation and dissemination, where does that leave “Us”?
REFERENCES


