
Can We Think Without Categories?

Lev Manovich

Abstract

In this article methods developed for the purpose of what I call “Media Analytics” are contextualized, put into a historical framework and discussed in regard to their relevance for “Cultural Analytics”. Large-scale analysis of media and interactions enable NGOs, small and big businesses, scientific research and civic media to create insight and information on various cultural phenomena. They provide quantitative analytical data about aspects of digital culture and are instrumental in designing procedural components for digital applications such as search, recommendations, and contextual advertising.

A survey on key texts and propositions from 1830 on until the present sketches the development of “Data Society’s Mind”. I propose that even though Cultural Analytics research uses dozens of algorithms, behind them there is a small number of fundamental paradigms. We can think them as types of data society’s and AI society’s cognition. The three most general paradigmatic approaches are data visualization, unsupervised machine learning, and supervised machine learning. I will discuss important challenges for Cultural Analytics research. Now that we have very large cultural data available, and our computers can do complex analysis quite quickly, how shall we look at culture? Do we only use computational methods to provide better answers to questions already established in the 19th and 20th century humanities paradigms, or do these methods allow fundamentally different new concepts?

Media Analytics and Cultural Analytics

Since the middle of the 2000s, global digital culture has entered a new stage that I call “media analytics.” (Manovich 2018: 473–488) Computational analysis of massive numbers of cultural artefacts, their online “lives,” and people’s interactions with these artefacts and each other has redefined dynamics and mechanisms of culture. Such analysis is now used by numerous players – the companies who run social networks, NGOs planning their outreach, millions of small businesses around the world advertising online, or millions of people who are using social media and web analytics tools and dashboards to refine their online posts and self-presentation and understand their social and professional value. For example, I am using Google Analytics

to understand how visitors interact with all pages on my websites. I can also look up my score on academia.edu, computed by comparing popularity of my publications on this network with publications of millions of other academics and students.

This large-scale analysis of media and interactions also enables other key components of digital culture such as search, recommendations, and contextual advertising. For example, to make its search service possible, Google continuously analyses full content and mark-up of billions of Web pages. It looks at *every* page on the Web that its spiders can reach – its text, layout, fonts used, images, and so on, using over 200 signals in total (Google, 2016a). E-mail spam detection relies on analysis of texts of numerous e-mails. Amazon analyses purchases of millions of its customers to recommend books. Netflix analyses choices of millions of subscribers to recommend films and TV shows. It also analyses information on all its offerings to create more than 70,000 genre categories (Madrigal, 2014). Contextual advertising systems such as AdSense analyse content of Web pages and automatically select the relevant ads to show. Video game companies capture gaming actions of millions of players and use this to optimize game design. Other examples include automatic translation and recommendations for people to follow or add to your friends list on social networks).

The same core algorithms used in the industry also make possible new research about cultures and societies in fields that include computer science, data science, computational social science, digital humanities, urban studies, media studies, data visualization, and data design. Since the research that uses large cultural datasets and data science to create, manage, and analyse them is spread between all these institutional and design disciplines, I have been using the umbrella term “Cultural Analytics” to refer to it. (Manovich 2009)

Here are a few examples of Cultural Analytics research: “Toward Automated Discovery of Artistic Influence,” (Saleh 2015) “Measuring the Evolution of Contemporary Western Popular Music,” (Serrà 2012) and “Shot durations, shot classes, and the increased pace of popular movies.” (Cutting/Candan 2015: 40–62) The first paper presents a mathematical model for automatic discovery of influence between artists. It then tests it using images of 1710 well-known paintings created by 66 well-known artists over a number of centuries. While some of the discovered influences are the same ones often described by art historians, the model also suggested other visual influences between artists not discussed previously. The second paper investigates changes in popular music using a dataset of 464,411 songs from 1955 to 2010. The dataset included “a variety of popular genres, including rock, pop, hip hop, metal, or electronic.” The authors concluded that over time, there was “the restriction of pitch transitions” and “the homogenization of the timbral palette” – in other words, some of the musical variability has decreased. The third paper analyses gradual changes in average shot duration across 9400 English-language narrative films created during 1912–2013.

In this article I will discuss a few general challenges for Cultural Analytics research. Now that we have very large cultural data available, and our computers

can do complex analysis quite quickly, how shall we look at culture? Do we only use computational methods to provide better answers to questions already established in the 19th and 20th century humanities paradigms, or do these methods allow fundamentally different new concepts?

I think that such perspectives are necessary because contemporary culture itself is now driven by the same or similar methods. And this is the key difference between using computational methods and concepts to analyse cultural data today vs. twenty years ago. Now these methods and concepts are driving everyday digital culture lived by billions of people. When small numbers of humanists and social scientists were analysing cultural data with computers in the second part of the 20th century, their contemporary culture was mostly analogue, physical, and non-quantifiable. But today we as academic researchers live in the “shadow” of a world of social networks, recommendations, apps, and interfaces that all use media analytics. As I already explained, I see *media analytics* as the new stage in the development of modern technological media. This stage is characterized by algorithmic large-scale analysis of media and user interactions and the use of the results in algorithmic decision making such as contextual advertising, recommendations, search, and other kinds of information retrieval, filtering of search results and user posts, document classification, plagiarism detection, video fingerprinting, content categorization of user photos, automatic news production etc.

And we are still only at the beginning of this stage. Given the trajectory of gradual automation of more and more functions in modern society using algorithms, I expect that production and customization of many forms of at least “commercial culture” (characterized by conventions, genre expectations, and templates) will also be gradually automated. So, in the future already developed digital distribution platforms and media analytics will be joined by the third part: *algorithmic media generation*. (Of course, experimental artists, designers, composers, and filmmakers have been using algorithms to generate work since the 1960s, but in the future, this is likely to become the new norm across culture industry.) We can see this at work already today in automatically generated news stories, online content written about topics suggested by algorithms, production of some television shows, and TV broadcasts during sport events where multiple robotic cameras automatically follow and zoom into dynamic human performances. So, for instance, if we want to analyse intentions, ideology and psychology of an author of certain cultural artefact or experience, this author maybe not a human but some form of AI that uses a combination of data analysis, machine learning and algorithmic generation.

Until ten years ago, key cultural techniques we used to *represent and reason* about the world and other humans included natural languages, lens-based photo and video imaging, various other media for preserving and accessing information, calculus, digital computers, and computer networks. The core concepts of *data/AI society* are now as important. *They form data society's “mind” – the particular ways of encountering, understanding, and acting on the world and the humans.* And this is

why even if you have no intention of doing practical Cultural Analytics research yourself, you need anyway to become familiar with these new *data-centred cultural techniques*. (The concept of “cultural techniques” has been mostly used in recent German media theory. See Winthrop-Young/Irascu/Jussi Parikka 2013).

While both media analytics in industry and Cultural Analytics research use dozens of algorithms, behind them there is a small number of fundamental paradigms. We can think them as types of data/AI society’s cognition. The three most general ones are *data visualization*, *unsupervised machine learning*, and *supervised machine learning*. Others are feature extraction, clustering, dimension reduction, classification, regression, network science, time series analysis, and information retrieval. (Others may have a different list depending on their field).

The Development of Data Society’s Mind: Short Timeline

Given that the terms “AI,” “machine learning” and “data science” have entered public discourse in 2010s, many people may think that these fields are very new. In fact, almost all ideas and methods used today for data analysis were developed in the 19th and 20th century. The following is my timeline of selected concepts and methods in statistics and computational data analysis. (What is included and excluded is based on my own experience with data science and is consciously biased towards Cultural Analytics research rather than all industry applications.)

The nineteenth century:

1. *Normal distribution* observed in physical characteristics of groups of people: Adolphe Quetelet in the early 1830s.
2. *Social physics*, 1835: Quetelet’s book *Sur l’homme et le développement de ses facultés, ou Essai de physique sociale* (translated in English as *Treatise on Man*) outlines the idea of “social physics.” Quetelet appropriates this term from Comte who was already using it earlier.
3. *Sociology*, 1838: Learning that Quetelet uses his concept, Comte introduces the new concept of “*sociologie*.”
4. *Standard deviation*: Francis Galton in late 1860.
5. *Correlation*: Aguste Bravais in 1846; rediscovered by Galton in 1888.
6. *Regression analysis*: Galton in the 1880s.
7. *Histogram*: Pearson in 1895.
8. Regression analysis, extended and formalized – Udney Yule and Karl Pearson, 1897–1903.
9. *Statistical Hypothesis Test theory* foundations: Pearson in 1900.
10. *Power Law*, 1896: Economist Vilfredo Pareto publishes his observations of “80–20 rule” that is latter named “Pareto Law”; it exemplifies more general “power law” that describes many phenomena in network culture and social media today such as “long tail.”

The twentieth century:

1. *Principal Component Analysis*: Pearson in 1901.
2. *Factor analysis*: Charles Spearman in 1904.
3. *Multi-variable regression* – first used the 1900s.
4. *Markov's Chains*, 1906: Andrew Markov starts working on theory and methods for statistical analysis of time series which are later names "Markov's chains."
5. *Network analysis*, the 1930s: Jacob Moreno develops methods for representing social networks; in the 1950s measurement methods and concepts for network analysis are formalized.
6. *Neural networks*, 1957: Frank Rosenblatt develops neural networks for classification ("perceptrons"). This research was based on the earliest model of neural networks developed by Ukrainian-born American scientist Nicolas Rashevsky. His students Walter Pitts and Warren McCulloch further developed his model and published their famous 1943 article in Rashevsky's journal. (McCulloch/Pitts 1943: 115–133)
7. "Software," 1958: John Tukey is the first to use this term. (Tukey 1958: 1–9)
8. *Multi-dimensional Scaling* (MDS), the 1950s. (Torgerson 1958; Green 1975: 24–31)
9. *Support Vector Machines* (SVM), 1963: Vladimir Vapnik and Alexey Chervonenkis (Soviet Union) develop the original algorithm. A version of the algorithm published by Vapnik in 1995 becomes one of the most popular algorithms for classification.
10. *Deep Learning*, 1965: Alexey Ivakhnenko and V.G. Lapa publish the first working algorithm for deep learning. In the late 2000s deep learning becomes most popular approach to classification. Ivakhnenko is often referred as the "father of deep learning."
11. *Vector space* concept, late 1960s: developed for text comparisons by Gary Salton.
12. *Exploratory data analysis*, 1971: Tukey starts developing ideas of exploratory data analysis.
13. *Statistical programming languages*, 1976: The language called "S" is developed in Bell Labs by John Chambers and others. It makes Exploratory Data Analysis possible since it allows researchers to perform every kind of statistical and data analysis from a command line and make graphs. (Chambers 2016) "R" language developed later from "S."
14. *Backpropagation*, 1986: This new method for teaching deep networking starts to become popular after the publication of the article by David Rumelhart, Geoffrey Hinton, and Ronald Williams in *Nature*. (Rumelhart et al. 1986: 533)
15. *PageRank*, 1996: Larry Page and Sergey Brin developed the PageRank algorithm. The idea was proposed earlier and can be traced back to the work of Eugene Garfield, the founder of bibliometrics and scientometrics in the 1950s. (Although PageRank is not the key algorithm in data analysis, it had fundamental effect on how people interact with information and cultural content online.)

16. *Item-item collaborative filtering*, 1998: Amazon invents and starts using Item-item collaborative filtering, a recommendation algorithm that calculates similarity between objects (such as books) based on people's ratings of those objects. (Sarwar 2001: 285–295)

The twenty-first century:

1. *Latent Dirichlet allocation* (LDA), 2003: The model for “topic modelling” in text documents called Latent Dirichlet allocation is published by David Blei, Andrew Ng, and Michael I. Jordan. J. K. Pritchard, M. Stephens, and P. Donnelly published the original topic modelling paper in 1999.
2. *Deep learning takes off*, 2006: The current wave of deep learning research starts with publication “A Fast Learning Algorithm for Deep Belief Nets” by Geoffrey Hinton, Simon Osindero, and Yee-Whye Teh. (2006: 1527–54) Combination of bigger data available for training, the use of computer clusters and GPUs, and sharing of datasets and code online, and competition such as ImageNet lead to very fast progress.
3. *Use of deep learning for image classification*, 2012: A paper “ImageNet Classification with Deep Convolutional Neural Networks” by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton shows that “deep learning outperforms other methods for image classification.”

Summarizing this timeline, I can say that practically all concepts and methods relevant for computational analysis of culture on large scale were invented by already the middle of the 1970s, although their massive applications in cultural industry and in Cultural Analytics research is only about ten years old.

Do We Want to “Explain” Culture?

Approaching cultural processes and artefacts as “data” can lead us to ask the kinds of questions about culture that people who professionally write about it, curate and manage it do not normally ask today – because such questions would go against the accepted understanding of culture, creativity, aesthetics, and taste in humanities, popular media or art world. For example, would collectors and museums pay millions for the works of contemporary artists if data analysis shows that they are completely unoriginal despite their high financial valuations? Or, if data analysis reveals that trends in art world can be predicted as accurately as the ones in fashion?

The most well-known and influential quantitative analysis of cultural data within social sciences remains Pierre Bourdieu's *Distinction* (1979). The data used in this analysis comes from the surveys of French public. For analysis and visualization of this data, Bourdieu used recently developed method of *correspondence analysis*. It is similar to PCA but works for discrete categories, showing their

relations in graphical form. For Bourdieu, this form of data analysis visualization went along with his theoretical concepts about society and culture, and that's why it plays a central role in this book. *Distinction* is Bourdieu's most well known book, and in 2012 Bourdieu was the second most quoted author in the world in academic publications, just behind Michel Foucault. (Truong/Weill 2012)

Bourdieu did not use the most common method of quantitative social science – “explaining” some observed phenomena using mathematic models such as linear regression. However, given the variety and the scale of cultural data available today, maybe today such method can produce interesting results?

What would happen if we also take other standard methods of quantitative social science and use them to “explain” the seemingly elusive, subjective, and irrational world of culture? For example, we can use factor analysis to analyse choices and preferences of local audiences around the world for music videos from many countries to understand the dimensions people use to compare musicians and songs. Or we can use regression analysis and combination of demographic, social, and economic variables to model choices made by “cultural omnivores” – people who like cultural offerings associated with both elite and popular taste. (Peterson 1992: 243–258)

In quantitative marketing and advertising research, investigators ask similar questions all the time in relation to consumer goods and cultural artefacts. And computer scientists also do this when they analyse social media and web data. But this does not happen in humanities. In fact, if you are in the arts or humanities, such ideas may make you feel really uncomfortable. And this is precisely why we should explore them.

The point of any application of quantitative or computational methods to analysis of culture is not whether it ends up being successful or not (unless you are in media analytics business). It can force us to look at the subject matter in new ways, to become explicit about our assumptions, and to precisely define our concepts and the dimensions we want to study.

So at least as a thought experiment, let's apply quantitative social science paradigm to culture. Quantitative social science aims to provide “*explanations of social phenomena expressed as mathematical relations between small numbers of variables* (what influences what and by how much). Once such models are created, they are often used for prediction. The common statistical methods for such “explanations” are regression models, versions of factor analysis or fitting a probability distribution to the data. The latter means determining if observed data can be described using a simple mathematic model, e. g., Gaussian distribution, log-normal distribution, the Pareto distribution etc. (In quantitative film studies, a number of researchers found that shot frequencies in the twentieth-century Hollywood films follow a log-normal distribution. See, for example, DeLong 2015: 129–36)

Are we interested in trying to explain what influences what in culture or predicting its future with mathematic models? Do we need to explain culture

through external economic and social variables? Do we really need to find that author's biography, for example, accounts for 30% of "variability" in her works? Or that age, location, and gender variables account for, let's say, 20% of variability in Instagram posts? And even if we find that a combination of some variables can predict the content and style of Instagram posts of some users with 95% accuracy, probably what is really important in this cultural sphere is the 5% we cannot predict.

Applied to real life data, regression models typically can only predict some of the data but not all of it. The part that is not predicted is often treated as "noise" because it does not fit the mathematical model. In fact, in the standard presentation of the regression analysis, the term that is added to the model to represent the unpredicted data is called an *error term*, or *noise*. The assumption is that the noise part is due to some possibly random variation, which adds disturbance to the process we are observing and modelling. However, this "noise" part is maybe most important in the case of cultural artefacts and experiences.

Is the Goal of Cultural Analytics to Study Patterns? (Yes and No)

Now that we understand the implications of looking at culture the way twentieth-century social scientists looked at society, do we actually want to do this? In Cultural Analytics we do not want to "explain" most or even some of the data using a simple mathematical model and treat the rest as "error" or "noise" just because our mathematical model cannot account for it. And we do not want to assume that cultural variation is a deviation from a mean. We also do not want to assume that large proportions of works in particular medium of genre follow a single or only a few patterns such as "hero's journey," "golden ratio" or "binary oppositions," or that every culture goes through the same three or five stages of development as it was claimed by some art historians in the nineteenth century.

I believe that *we should study cultural diversity without assuming that it is caused by variations from some types or structures*. This is very different from modern thinking of quantitative social science and statistical thinking it adapted. The historical development of statistics in the eighteenth and nineteenth century leads it to consider observed data in terms of deviations from the mean.

Does this mean that we are only interested in the differences and that we want to avoid any kind of reduction at all cost? To postulate existence of cultural patterns is to accept that we are doing at least some reduction when we think and analyse data. Without this, we cannot compare anything, unless we are dealing with extreme cultural minimalism or seriality, where the artist makes everything else equal and only varies a single variable, like Sol LeWitt or some minimalist music.

Cultural Analytics thus can be also defined as *the quantitative analysis of cultural patterns on different scales.*" But if we accept this definition, we need to

immediately add an important point to this statement. *While we want to discover repeating patterns on different scales in cultural data, we should always remember that they only account for some aspects of the artefacts and their reception.*

Unless it is a 100% copy of another cultural artefact or produced mechanically or algorithmically to be identical with others, every expression and interaction is unique. In some cases, this uniqueness is not important in analysis, and in other cases it is. For example, the face features we extracted from a dataset of Instagram self-portraits revealed interesting differences in how people represent themselves in this medium in particular cities and periods we analysed. But the reason we do not get tired looking at endless faces and bodies when we browse Instagram is that each of them is unique.

The key goal of Cultural Analytics as I see it should be to map in detail (and help us to think about these maps) the *diversity* of contemporary professional and user-generated artefacts created globally – i. e. to focus on what is different between numerous artefacts and not only on what they share. In the nineteenth and twentieth century the lack of appropriate technologies to store, organize, and compare large cultural datasets was contributing to the popularity of reductive cultural theories. Today I can use any computer to map and visualize thousands of differences between tens of millions of objects. We do not have an excuse any more to only focus what cultural artefacts or behaviours share, which is what we do then we categorize them, or perceive them as instances of general types. So while we may have to start with extracting patterns first just to draw our initial maps of contemporary cultural production and dynamics given its scale, eventually they may recede in the background or even completely dissolve, as we focus only on the differences between individual objects.

Can We Think (About Culture) Without Categories?

In my experience, these ideals are easier to state than to put in practice. Human brain and languages are categorizing machines. Our cognition constantly processes sensory information and categorizes it. A pattern we observe is like constructing a new category: a recognition that some things or some aspects of these things have something in common. Can we learn to think about culture without categories?

How do we move away from the assumption of humanities (that until now “owed” thinking and writing about culture) that their goal of research is discovery and interpretation of general cultural types, be they “modernism,” “narrative structures,” “selfies,” or “amateur digital photographers”? How do we instead learn to see cultures in more details, without immediately looking for, and noticing, only types, structures or patterns?

Of course, first we need sufficiently large data samples, or ideally all artefacts. Next, we need to extract sufficiently large numbers of features that capture characteristics of these artefacts, their reception and use by audiences, and their

circulation. (We also need to think more systematically about how to represent cultural processes and interactions – especially since today we use interactive digital cultural media as opposed to historical static artifacts.) Once we have such datasets, we can explore them using various visualization techniques that work with results of exploratory data analysis – visualizations of cluster analysis, visualizations of distance matrixes, and visualizations that use dimension reduction (PCA, t-sne, etc.).

But what are the goals of these “explorations”? What are we measuring and comparing? As a way of conclusion, I want to propose one answer to this question. To observe and analyse culture means to be able to map and measure three fundamental characteristics. These characteristics are *diversity*, *structures* (e. g., clusters networks, and other types of relations), and *dynamics* (temporal changes). In the case of cultural situations where we may expect many works to follow some prescriptive aesthetics or use templates – for example, Instagram filters provided by the app, or the themes described and illustrated in thousands of advice posts – we can also look at the forth characteristic: *variability*. So, for example if we analyse a sample of Instagram images, we can first detect the presence of the themes that appear in many posts, and then look at deviations from these themes and also images that do not follow any of them. But we do not want to assume that the deviation from the type (or from a mean or another statistic we can compute for our dataset) is a necessary measurement for all cultural situations. The development of the appropriate *measures of cultural diversity, structure, dynamics and variability for different types of media and cultural fields* is itself a big theoretical and practical task. I see this as the central task for Cultural Analytics in years to come.

References

- Chambers, John M. (2016): *Stages in the Evolution of S*, March 7, 2000, accessed August 29, 2016, <http://ect.bell-labs.com/sl/S/history.html>
- Cutting, James E. and Ayse Candan (2015): Shot durations, shot classes, and the increased pace of popular movies. In: *Projections: The journal for movies and mind* 9(2): pp. 40–62, <http://people.psych.cornell.edu/~jec7/pubs/cutting&candanProj15.pdf>
- DeLong, Jordan (2015): Horseshoes, Handgrenades, and Model Fitting: The Log-normal Distribution Is a Pretty Good Model for Shot-length Distribution of Hollywood Films. In: *Digital Scholarship in the Humanities* 30/1, pp. 129–36.
- Green, Paul E. (1975): Marketing Applications of MDS: Assessment and Outlook. In: *Journal of Marketing* 39/1 (1975): pp. 24–31.
- Hinton, Geoffrey, Simon Osindero, and Yee-Whye Teh (2006) “A Fast Learning Algorithm for Deep Belief Nets,” *Neural Computation* 18, no. 7 (2006): 1527–54
- Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton (2012): ImageNet Classification with Deep Convolutional Neural Networks. In: *Advances in Neural*

- Information Processing Systems 25* (2012), accessed August 8, 2016, <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks>
- Kurenkov, Andrey (2016): A 'Brief' History of Neural Nets and Machine Learning, December 24, 2015, accessed August 29, 2016, <http://www.andreykurenkov.com/writing/a-brief-history-of-neural-nets-and-deep-learning-part-4/>
- Manovich, Lev (2018): 100 Billion Data Rows per Second: Media Analytics in the Early 21st Century. In: *International Journal of Communication* 12 (2018): pp. 473–488.
- Manovich, Lev (2009): *Cultural Analytics: Visualizing Cultural Patterns in the Era of "More Media"*. In: *Domus* (March 2009).
- McCulloch, Warren S. and Walter Pitts (1943): A Logical Calculus of the Ideas Immanent in Nervous Activity. In: *The Bulletin of Mathematical Biophysics* 5/4, pp. 115–133, doi: 10.1007/BF02478259
- Rumelhart, David, Geoffrey Hinton, and Ronald Williams (1986): Learning Representations by Back-propagating Errors. In: *Nature* 323/6088: 533
- Saleh, Babak et al. (2015): Toward Automated Discovery of Artistic Influence, <https://arxiv.org/abs/1408.3218>
- Sarwar, Badrul et al. (2001): Item-based collaborative filtering recommendation algorithms. In: *Proceedings of the 10th International Conference on World Wide Web* (ACM): 285–295.
- Serrà, Joan et al. (2012): Measuring the Evolution of Contemporary Western Popular Music, 2012, <https://www.nature.com/articles/sre00521>
- Torgerson, Warren S. (1958): *Theory and Methods of Scaling*, New York: Wiley
- Truong, Nicolas and Nicolas Weill (2012): A Decade after His Death, French Sociologist Pierre Bourdieu Stands Tall. In: *The Guardian*, February 21, 2012, <https://www.theguardian.com/world/2012/feb/21/pierre-bourdieu-philosophy-most-quoted>
- Tukey, John W. (1958): The Teaching of Concrete Mathematics. In: *The American Mathematical Monthly* 65/1: pp. 1–9.
- Winthrop-Young, Geoffrey, Ilinca Irascu and Jussi Parikka (2013): Cultural Techniques (Special Issue), *Theory, Culture & Society* 30/6 (November)

