Culture Analytics: An Introduction

James Abello (Rutgers Univ and DIMACS)
Katy Börner (Indiana Univ)
Russ Caflisch (UCLA)
Jianbo Gao (Guangxi Univ)
Lev Manovitch (CUNY Graduate Center and UCSD)
Tina Eliassi-Rad (Northeastern Univ)
Vwani Roychowdhury (UCLA)
Timothy R. Tangherlini (UCLA)

Introduction

The data driven analysis of culture is now a reality, and calls for substantial investment in the emerging field of Culture Analytics.[1] Culture Analytics simultaneously studies the dynamics of culturally informed interactions between people, and the cultural expressive forms that result from these interactions, and it does so at scales hitherto unimaginable. Researchers from the Humanities, the Social Sciences, the Mathematical Sciences, and the Data Sciences are now collaborating to identify, document, and integrate concepts, methods and tools that will provide an intellectually and ethically sound approach to the study of cultures across time and across space, leveraging the enormous gains made in the past decade in computation and machine readable cultural archives, from libraries and museum collections to the born digital cultural expressions of billions of people on the internet. This rapid proliferation of digital data has made the role of Culture Analytics all the more central particularly given the potential for significant benefit that lies in harnessing the domain expertise of researchers across these disciplines. The fundamental need for a collaborative approach that traverses disciplines which traditionally have little interaction will ensure that Culture Analytics matures rapidly, and that data-driven studies of culture will be both responsible and meaningful.

Culture Analytics is, by definition, a collaborative, translational data science that explores culture and cultural interaction as a multi-scale / multi-resolution phenomenon. The macroscopic view, that allows a researcher to move from the microscale of close reading, up through the mesoscales, and on to the macroscale of distant reading, is a hallmark of the discipline. Disciplines that are traditionally focused on the study of culture from the Humanities and Social Sciences are central in the development of research questions and techniques, and scholars from these fields working collaboratively with Engineers, Computer Scientists and Applied

[1] We deliberately use the term “Culture Analytics” as opposed to the term “Cultural Analytics” (Manovich 2007). In this formulation, “Culture” as a noun, is the object of analysis--an analysis that recognizes that culture is a dynamic process and, through that process, cultural artifacts--expressive forms that are remarkably efficient at encoding cultural ideology (norms, beliefs, values)--are created. Thus, culture analytics is the data-driven analysis of culture.

Mathematicians can ensure that none of the work is conceptualized as service. Rather, we recognize that the teams that work best together have the best chance at unraveling some of the extraordinary complexity of culture.

A main concern of Culture Analytics researchers is how best to formalize the validity of cultural hypotheses and models obtained via the analysis of organic data either collected passively or collected for other purposes. This challenge is particularly crucial given the enormous amount of effort directed at the analysis of cultural phenomena from sources such as the internet, where controls are not put in place by design. While the Social and Behavioral Sciences are used to generating data from experiments that are carefully designed, and while the Humanities rely on rich historical archives of textual and material representations of past cultural activity, these scholarly traditions may need to be updated in the context of the explosion of digital resources that reflect cultural processes. Some researchers in these fields are, and perhaps rightly so, skeptical about any claims made from large-scale organic cultural data, yet it is precisely this type of data that fuels a great deal of investigation in the data-driven study of culture as currently practiced. Consequently, there is room for considerable formal work to establish metrics for evaluating when a data set can be trusted. We consider this one of the central open challenges for Culture Analytics. As an intermediate step, we propose that the generation of hypotheses and their verification via controlled experiments may provide opportunities to move incrementally toward meeting the challenges related to data integrity.

The ethics of cultural data analysis is similarly a central concern of Culture Analytics. The data-driven analysis of culture can lead to the discovery of both socially constructive and socially destructive patterns. Similarly, there is a strong possibility that individuals' privacy as well as safety can be at risk. Recognition of these risks, while also recognizing the massive social and cultural benefits that can derive from cultural analysis at scale are crucial to the success of the discipline. While civil unrest and wars have threatened cultural heritage sites throughout North Africa and the Middle East, techniques related to rapid crowd-sourced documentation, as well as the prediction of at-risk sites, can be deployed as a means for attenuating these threats. Similarly, culturally sensitive responses to natural disasters could allay some of the profound missteps of earlier responses such as those in the aftermath of Hurricane Katrina. Algorithms designed to work with large scale cultural data need to be sensitive to the biases that may creep into this work. For instance, in a well known failure, Google Photos accidentally labeled photographs of black people as gorillas, a deeply offensive, regrettable, and avoidable situation.[2] Understanding cultural processes, and providing robust documentation of cultural artifacts, all within an ethical context sensitive to an individual's human rights is a core aspect of Culture Analytics.

As a first step toward articulating the goals and challenges of Culture Analytics, NSF's Institute for Pure and Applied Mathematics (IPAM) at UCLA sponsored a long program during the spring 2016. The program, organized into a weeklong tutorial, four weeklong workshops, and a concluding workshop at Lake Arrowhead, brought together over 150 scholars as participants and speakers to address these complex issues. The goal of the long program was

[2] Alistair Barr, "Google Mistakenly Tags Black People as 'Gorillas,' Showing Limits of Algorithms", The Wall Street Journal, July 1, 2015. http://on.wsj.com/1CaCNlb

to offer an inflection point in the development of the field, recognizing the acceleration of work on cultural problems across many disciplines that take advantage of the explosion in machine-actionable cultural data, from historical archives of texts to rich 3D representations of archeological sites, from Twitter commentary on current events to huge collections of user interaction data on dating sites. By coming together and learning across disciplines, the goal is to move toward a collaborative, translational data science for Culture Analytics that converges on a shared understanding of the challenges associated the data driven analysis of cultural phenomena.

The Arrowhead Problems in Culture Analytics

Inspired by the now classic “Hilbert Problems” in mathematics (1900), the participants in the Culture Analytics long program devised a series of thirteen, Hilbert-like problems, that may provide clear goals for the development of Culture Analytics in the years to come.

We explain briefly below each problem what we understand to be the core of the problem, and why we believe it to be one of the essential problems in the discipline. Many of the problems have no clear resolution at the current time, and several problems may be unresolvable. We do not expect these to be easy or resolved in the near future, but we do find them to be of such fundamental nature, that even partial solutions will help drive the field of Culture Analytics in the years to come.

1. Metrics for the study of culture(s)

2. Identifying, defining and measuring cultural complexities

3. Is culture automatically the result of the evolution of groups (see Hilbert Problem #5)

4. What are the fundamental mechanisms for cultural network formation?

5. Find algorithms to detect the culturally meaningful topical structure of heterogeneous cultural data (see Hilbert Problem #10)

6. Find algorithms to detect culturally meaningful phase transitions in heterogeneous cultural data (see Hilbert Problem #10)

   By analogy with transitions between solids, liquids, and gaseous states of matter, a cultural phase transition is conceptualized as a change in a feature of a cultural system. These phase transitions include emergent large-scale ordered or disruptive cultural movements, as well as the formation and change of large-scale social opinion or mood. These transitions are often accompanied by large changes in cultural entropy, and thus in principle can be readily detected.

7. Identify invariance of offline and online culture(s) to understand their co-evolution

8. Measure the impact of culture on social conflict, inequality, health (e.g., there are different narratives/reasons why groups of people do not vaccinate their kids), and the environment. Disentangle correlation and causation, aim to understand delays (e.g., the time it takes between a new culture becoming available and its impact on daily life) and positive plus negative feedback cycles.

9. Measure the rate of change in culture, identify what influences this change, and map the the interdependencies between online and offline cultures.
   Culture is an ever changing process. How should we measure this change? What does the rate of change tell us about the robustness of a culture? What internal and external, online and offline forces cause change in a culture?

10. Scaling algorithms to any heterogeneous cultural data

11. Can one develop a calculus of culture?

12. Are there axioms of culture and can one develop a mathematical treatment of these? (see Hilbert Problem #6)

13. How do we measure the cultural impact of globalization?
    A recent study by Facebook Research reported the following: "Each person in the world (at least among the 1.59 billion people active on Facebook) is connected to every other person by an average of three and a half other people."[3] As people from various cultures come together in forums such as Facebook, creating new virtual cultural groups, nuances of the individual cultural backgrounds of the members of these new groups change through processes such as hybridization, creolization or other negotiated forms of in-group cultural alignment. This globalizing tendency toward cultural alignments has advantages and disadvantages. For example, humans are inherently tribal; how does this trait manifest itself in these globalized virtual communities?

Grand Challenges for Culture Analytics

[3] https://research.facebook.com/blog/three-and-a-half-degrees-of-separation/

Along with the Arrowhead Problems in Culture Analytics, the participants in the IPAM long program identified a group of Grand Challenge areas for Culture Analytics. It is in these areas that we believe the greatest gains can be made, particularly by groups working with the processes and approaches that we discuss in the six Culture Analytics white papers. These Grand Challenge areas are broken into four main groupings: Theoretical Challenges, Empirical Challenges, Applied Mathematics/Computer Science Challenges, and Practical Societal Challenges. We deliberately do not specify a Humanities and/or Social Science set of challenges as all of these challenge areas can be seen as subsumed under the larger umbrella of the study of culture at scale.

We propose to curate these Grand Challenges for Culture Analytics on a wiki page, allowing for CA researchers to propose challenges and sub-challenges in the coming years.

- Theoretical Challenges
  - Properties of cultural systems (what correlations, statistical models, laws exist?)
  - Phase transitions (e.g., the shifting perception of tattoos over time and across space)
- Empirical Challenges
    - Hypothesis testing of cultural assumption (e.g., acceleration of culture, the perception of time compressing as time progresses)
    - Validate humanistic approaches for understanding culture
    - Validate cultural paradoxes
    - A/B testing in cultural systems
    - Creation of an open source software package that is accessible, module, adaptable and that allows for reproducibility in Culture Analytics research
- Engineering/Computer Science/Applied Mathematics Challenges
  - Scalability (to trillions of records/PB of data)
  - Usability
  - Interpretability
  - Reproducibility
  - Long tail distributions for cultural phenomena
- Practical/Societal Challenges
  - Improve health (e.g. Use social media data to predict and prevent episodes of depression)
  - Measure, model, and promote cultural diversity
  - Reduce substance abuse and identify communities at risk (e.g. Contextualize people's experiences and behaviors in overall contexts)
  - Promote global peace by analyzing media reported events and identifying bifurcation points
  - Promote social stability by decreasing inequalities in education, economics, and health
  - Reduce education inequality and increase the teaching of diverse cultures
  - Privacy
  - Ethics
  - Access

Conclusion

Culture Analytics is moving out of its infancy, but it is still a young discipline. Drawing on the groundbreaking work of colleagues in Digital Humanities, Computational Social Sciences, Data Science, Computer Science, Mathematics, and Engineering, Culture Analytics neither subsumes nor is subsumed by any of these disciplines. Instead, it stands as an exciting new translational data science, one that takes advantage of the expertise of colleagues across the cultural disciplines, from Humanities, with its focus on cultural artifacts, to Social Sciences, with its focus on cultural processes, from Library Science, with its focus on archiving of cultural expressive forms. These researchers are wedded in complex teams with colleagues from the Mathematical, Engineering and Computer Sciences, all with the goal of unraveling the remarkable complexity of culture and promoting positive social developments.

Additional White Papers in the Culture Analytics Series:

1. Cultural Archives: Enhancing Materials, Collections, Formats, Data for Culture Analytics
2. Collaboration and Translational Data Science for the Study of Culture
3. Interfaces and User Experience Design for Culture Analytics
4. Cultural Phenomena at Multiscale and Multiresolution: Powers and Perils
5. Language and Culture Analytics
6. Teaching Culture Analytics