Towards Critical Deployment of Embedded Humanities in Computational Analysis

Atilio Barreda II
Professor Rhody
Feminist Text Analysis
26 May 2023

Contents

Abstract

This suggestion paper embarks on a critical exploration of computational and algorithmic analysis methods, particularly their roots in phallogocentric including textual analysis. paradigms, thus pitching universalist, omniscient perspectives. It seeks to resist the allure of a 'magic button' (Leurs) or an inherently feminist algorithm, concepts that promise a straightforward fix of these fundamentally embedded biases. Instead, it advocates for careful navigation through these socio-technical challenges and assumptions, thereby contesting simplistic narratives of technological determinism and foregrounding the importance of critical and close reading of computional analysis methods.

Drawing from the works of Sir Francis Galton as an illuminating case study, the research examines the sexist biases evident in the development of correlation and regression - foundational models of statistical and computational methodologies. By engaging with Donna Haraway's notion of 'situated knowledges', it emphasizes that knowledge is partial and deeply entwined with particular humanistic, ethical, and cultural contexts.

The study further mobilizes transfer learning – an application of adapting an already trained machine learning model to new tasks – to illustrate Haraway's idea of situated knowledge within the sphere of machine learning. This approach challenges the supposed universality of machine learning models and emphasizes the potential of contextual adaptation in these models. Whether through fine-tuning the model for specific tasks or interpreting fresh data through the framework of existing models, this paper underscores the context-dependent nature of epistemic outcomes.

By advocating for critically deploying these computational techniques, the paper engages in a call for *continual* deconstruction of these methods. Drawing from critical theory can unpack whether analysis methods' philosophical and humanistic assumptions align with the task or dataset at hand. This approach facilitates a nuanced understanding of the dynamic interaction between computational methodologies and the humanities, contributing to the larger discourse on their integration and mutual reconfiguration.

Atilio Barreda II

Professor Rhody

Feminist Text Analysis

26 May 2023

Towards Critical Deployment of Embedded Humanities in Computational Analysis

> "She would not say of any one in the world now that they were this or were that...she would not say of herself, I am this, I am that."
>
> — Mrs. Dalloway (Woolf)

1. Introduction

Computational methodologies have radically altered our engagement with and perception of information and are becoming ubiquitous in daily life as well as in humanities research. While opening up new opportunities and potential, this surge also presents challenges within the humanities and social sciences. As scholars across diverse disciplines seek to incorporate these computational tools into their studies, a critical question arises: how can we ensure these methodologies do not unwittingly perpetuate harmful biases and structures inherent in their initial design, especially when rooted in imperialist and sexist paradigms? *Can there be a feminist text analysis?*

In this paper I navigate these critical intersections of computing and humanities, particularly focusing on foundational statistical models and Transfer Learning (TL). TL is a machine learning approach where knowledge gleaned from solving one problem is utilized to address different but related problems. Critical engagement with computational tools has implications that go beyond performance and accuracy.

I propose to view these intersections through the lens of feminist analysis and Donna Haraway's conception of 'situated knowledges.' Haraway argues that knowledge is neither static nor universally applicable but is partial and situated, deeply rooted within specific socio-technical contexts. By recognizing this, we can critique and navigate the entanglements of humanities and computational techniques more effectively. This departure from the universalist narratives enables us to examine and comprehend the embedded humanities within computational methods like correlation, regression, and transfer learning more fully, and apply these methods with an informed, critical lens. I will also leverage Derrida's concept of 'phallogocentrism[1],' which challenges the privileging of masculine over feminine, speech over writing, and presence over absence within Western philosophical thought. This interdisciplinary exploration aims at fostering a nuanced understanding of analysis methods from a humanistic, reflexive standpoint, promoting the cultivation of ant-imperialist and feminist data analysis practices.

2. The Imperative of Critical Theory in Computational Methodologies: A Feminist Perspective

The intersection of computional methods and humanities forms a productive space of exploration, particularly when considering the historical and socially-guided development and foundational models of statistics. In particular, Galton's early work on correlation and regression provides an illuminating case study. Galton, a Victorian eugenicist and statistician, is perhaps most known for his development of the concept of regression to the mean and his studies on heredity. In his initial studies on human height, Galton assumed that a son's height was solely influenced by the father's height (**263**; Ethington et al.), a claim that is congruent with patriarchal philosophies and customs prevalent during his time.

This assumption mirrored the prevailing societal notions of patrilineal descent and the primacy of masculine influence, as well as the absurd notion that women served merely as physical vessels for masculine agency or presence. By viewing agency as a unified object and attributing its origin to fathers, Galton projected social structure onto the realm of biological heredity, exemplifying the intertwined nature of science and culture.

However, later studies proved Galton's hypothesis incorrect. Both parents, not just fathers, influence a child's height. These studies also align with feminist critiques of cultural models that have gender hierarchy as a conceptual and ethical base. However, Galton's initial hypothesis serves as evidence for the biases and societal structures that can seep into the 'objective' domain of scientific research.

[1] The term 'phallogocentrism' is a neologism developed by Jacques Derrida. It is a combination of 'phallocentrism' and 'logocentrism'. Phallocentrism represents the privileging of the male or the masculine in interpreting meaning, culture, or *text*. Logocentrism, denotes the privileging of speech over writing in Western thought which serves as a signifier of a belief in a perfect, original meaning that precedes and transcends language. Derrida coined the term 'phallogocentrism' to critique these dual tendencies in Western philosophical thought that privilege the masculine and the supposed 'primary and absolutely irreducible signified' (**23**; Derrida) in speech acts. This double privileging constitutes an exclusionary binary that marginalizes other forms of expression, such as those that are feminist interpretations, 'derived' speech , or otherwise different from the naturalized, yet arbitrary norm. Phallogocentrism reinforces a metaphysical hierarchy that underlies and informs a range of social and cultural practices. (**lxix**; Derrida)

Instead of an objectivity based on an impossible detachment from social context and circumstance, Haraway advocates for a 'doctrine of embodied objectivity that accommodates paradoxical and critical feminist science project' **(581)**. Scientific objectivity is not a detached, universal perspective but a specific, situated knowledge, influenced by the biases and the context of its time. The patriarchal biases inherent in Galton's work, and their subsequent correction, demonstrate how scientific 'truths' are rooted in their specific historical and cultural contexts. In this light, computational methods and machine learning algorithms, like all scientific endeavors, should be understood and evaluated within the context of their creation and use, shedding light on the underlying biases that might influence their operation and outcomes.

Haraway references Katie King's observation that 'Rational knowledge is a process of ongoing critical interpretation among "fields" of interpreters and decoders. Rational knowledge is power-sensitive conversation' **(590)**. Thus, rational knowledge, instead of being an unchanging monolith-*cum*-manifold[2], is revealed as a dynamic discourse inextricably intertwined with power dynamics and societal structures. This underscores the crucial need to continually engage with critical theory to maintain the ongoing, power-sensitive conversation that is the hallmark of a feminist rational knowledge. Doing so ensures a continual close reading of underlying assumptions, biases, and power structures within scientific research and particularly within the burgeoning field of computational methods and machine learning algorithms, thereby fostering a more equitable, robust, and rational knowledge production process.

This paper's epigraph features the eponymous protagonist, Clarissa Dalloway, from *Mrs. Dalloway*, as she reflects on her identity and experience of the world. Here Woolf posits that fluidity and complexity are inherent to identity. Woolf, through Clarissa, rejects simple, discrete definitions of identity. This rejection of absolute definitions resonates profoundly with the deconstruction of historical underpinnings and assumptions present in analytical methodologies. The early work of Galton on correlation and regression provides a prime example, wherein he attempted to reduce the complexity of human heredity to a binary classification. His presumption rested on the humanistic assumptions and ethics of his time.

The same essence of fluidity and complexity must be embodied in our approach towards textual

[2] In machine learning, there's a whole field dedicated to 'manifold learning'. The goal here is to discover the structure of high-dimensional data by finding the manifold that the data lies on. The idea is that high-dimensional data points like word embeddings often lie on or near a lower-dimensional manifold. Uncovering this manifold is said to reveal the true structure of the data

analysis methods. Galton's initial hypothesis, later disproven, emphasizes that biases and societal structures can infiltrate even the most seemingly 'objective' domains of scientific research, including computational methods. As Leurs explains, this framework allows 'data scholars across the humanities and social sciences to produce more robust and meaningful stories rather than universal truths or disembodied generalisations' **(133)**.

Leurs further emphasizes the simultaneous promise and peril brought by the advent of big data, particularly its implications for women and minority populations. He suggests that our current data-driven research landscape, often propelled by a 'data-rush' mentality **(150)**, inadvertently reifies dominant ideologies, thereby correlating technological advancements with social progress. However, this emphasis on technology overlooks the power structures and biases embedded within these 'objective' data sets and analytical tools.

The concerns that Leurs brings to the forefront align with feminist and postcolonial critiques of mainstream science and highlight the urgency of integrating these perspectives into data studies. Conventional empiricist knowledge production often operates under the illusion of a detached, knowable world, where the researcher assumes a position of transcendent objectivity, or a 'god-trick' **(136)**, as Haraway *via* Leurs terms it.

Drawing on Donna Haraway's situated knowledges, the necessity of acknowledging that our knowledge, like the characters in Woolf's novel, are not static, but rather are deeply embedded in our social and technical context. It is only by embracing this complex, situated objectivity that we can begin to navigate the entanglements of the concerns of humanities and those of computational techniques. In doing so, we open up space for a critical, nuanced understanding of computational methodologies and their application within the humanities, underpinning the development of feminist and anti-imperialist data analysis practices.

3. Understanding Transfer Learning in Textual Analysis

3.1 What is Transfer Learning?

Transfer Learning is a powerful technique in the field of Machine Learning, which leverages preexisting knowledge from one context to accelerate or enrich learning in a different context. It's akin to applying skills or knowledge you've acquired in one domain (like learning a language) to a new, but related domain (like learning a different but related language). In the context of machine learning, transfer learning often involves using a pre-trained model, which has been trained on a large, diverse dataset, as a starting point for a new task.

3.2 Transfer Learning in Text Analysis

In text analysis, transfer learning has been revolutionary due to the varied and structured nature of language across different tasks. Models such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pretraining Transformer) are trained on a huge corpus of text. During this training, the models learn the internal structures of language, including grammar and context.

Once trained, these models can then be fine-tuned for specific tasks such as sentiment analysis, entity recognition (identifying important elements like names, places), or text classification (categorizing text into predefined categories).

4. Critical Use of Transfer Learning: An Application of Haraway's Situated Knowledges

$$\begin{tikzcd}
OldModel \arrow[r, "FeatureExtraction"] \arrow[d, "FineTuning"'] & NewData \arrow[d, "SituatedKnowledges"] \\
TransformedModel \arrow[r, "ModelValidation"] & FeministInsights
\end{tikzcd}$$

Figure 1: Proposal for feminist application of Transfer Learning

Transfer learning can serve as a practical example of Haraway's 'situated knowledges' concept in the context of machine learning. The diagram presented above illustrates this. The horizontal arrows denote a methodological transfer. Here, learned features from one problem (the *Old Model*) are applied to a new, related problem (*New Data*). Similarly, the *Transformed Model* uses its own new parameters and weights [3] to produce *Feminist Insights* with *Model Validation* geared toward feminist critique.

The vertical arrows in the diagram symbolize a generative process. The arrow labeled *Fine Tuning*

[3] In machine learning, a model learns from the data by adjusting its parameters, often termed 'weights'. These weights determine the importance assigned to each data set feature in making predictions. For example, in a machine learning model predicting real estate prices, the weight assigned to the 'size of the house' feature determines how much the size will impact the predicted price. When we talk about the model's 'new parameters and weights' being used to produce insights, we're essentially talking about how the model's learned importance of different features, its understanding of the data, is applied to generate these insights. Parameters and weights of a model are analogous to the features of a dataset. They represent what the model has learned about the importance and interactions of these features.

represents the adaptation of an existing analysis method to a new dataset. The result is the *Transformed Model* at the bottom left of the diagram, which is suited to a different, potentially previously marginalized task.

On the right side of the diagram, we have *New Data*, representing a new problem that we aim to solve using a previously underutilized dataset. The arrow marked *Situated Knowledges* depicts the acknowledgement of feminist critique into the data or problem space. This results in the production of *Feminist Insight*. *Model Validation*, the final step, tests the *Transformed Model* against insights from critical theory and feminist humanist interventions.

The original model is transformed into a more context-sensitive version, the *Transformed Model*, which accounts for the distinct nuances of the new problem or ignored perspective. The context-dependent nature of knowledge is recognized in the 'fine-tuning' process and subsequent *Model Validation* This emphasizes the importance of understanding each problem in its unique socio-technical context.

New data introduces new 'situated knowledges'. This indicates the need to critically engage with the data, question underlying assumptions, and explore power dynamics embedded in our models and algorithms. The feature extraction and transfer process is not inherently feminist. However, when combined with critical feminist analysis, it becomes a tool to reveal the *situatedness* and potential biases of machine learning models.

This diagram proposes a feminist approach to transfer learning, emphasizing that models are rooted in specific socio-technical contexts. The potential of transfer learning to adapt to new contexts is underscored, challenging the perceived universality of machine learning models. This approach supports the broader feminist and anti-imperialist critique of data analysis practices, stressing the need to question and disrupt embedded power dynamics in technology.

A critical application of transfer learning, alongside a feminist lens of analysis, can challenge and disrupt naturalized humanities ethics.

5. Conclusion

In this paper, we have discussed the intersections of computational methodologies and humanities through a critical perspective, how patriarchal biases are embedded into foundational statistical reasoning, tracing back to Galton's flawed hypothesis and revealing the inherent *situatedness* of scientific truths. We posited the necessity of adopting a situated and feminist objectivity.

In the realm of machine learning, we put forth an application of Haraway's 'situated knowledges' through the process of critique-informed Transfer Learning. We proposed a critical and cyclical process of continual fine-tuning, validation, and refining of models. The acknowledgement of feminist insight within this process serves to illuminate and challenge the biases present within our models and algorithms.

The methodology we suggest emphasizes the necessity of a continuous critical engagement, which brings forth a more just and rational process of knowledge production. It encourages the development of feminist and anti-imperialist data analysis practices. Our exploration underscores the need for an ongoing dialogue between the humanities and computationa analysis methods.

Works Cited

Derrida, Jacques. *Of Grammatology.* Translated by Gayatri Chakravorty Spivak, John Hopkins UP, 2003.

Ethington, Corinna A., et al. "Back to the Basics: Regression as It Should Be." *Higher Education: Handbook of Theory and Research*, edited by John C. Smart and William G. Tierney, Springer Netherlands, 2002, pp. 263–93, https://doi.org/10.1007/978-94-010-0245-5_6.

Leurs, Koen. "Feminist Data Studies: Using Digital Methods for Ethical, Reflexive and Situated Socio-Cultural Research." *Feminist Review*, vol. 115, no. 1, 2017, pp. 130–54. *https://doi.org/10.1057/s41305-017-0043-1*, https://doi.org/10.1057/s41305-017-0043-1.

Woolf, Virginia. *Mrs Dalloway.* Harcourt, Brace, 1925.