Deriving Architectural Inspiration from Big Data

Ethan Trapold and Karla Saldaña Ochoa

College of Design, Construction, and Planning, University of Florida

Faculty Mentor: Karla Saldaña Ochoa, School of Architecture

Abstract

It has been said that one of the most defining features of the 21st century so far has been the impact of social media. Trillions of data points exist in the digital realm, all charged with some sort of significance depending on what we are looking for. The proposition of this paper is to explore how this data can be used as a driver for architectural design, and how other digital tools along with artificial intelligence can work symbiotically with our own design intentions and tendencies. The research goal of this paper is to briefly review and analyze multiple studies and journal articles that utilize big data, social media, and artificial intelligence to address issues in the realm of architectural and urban design. From these studies, we will summarize the findings and synthesize the commonalities in their methods to speculate on how we could aggregate data to derive inspiration for an architectural design. Overall, this paper concludes that the methods for collecting and utilizing data across the field of design and urban planning can be adopted and utilized to collect data for the purpose of informing or inspiring an architectural design.

Keywords: Big Data, Artificial Intelligence, Social Media, Architectural Design

Introduction

At the very forefront of this paper is the concept of Big Data. The term Big Data refers to a data set that is extremely large and complex to the extent that traditional data processing methods cannot begin to analyze. It is important to note that Big Data as a term also encompasses the associated technologies that coincide with the data itself, such as machine learning and data mining. Data generated on social media platforms absolutely qualifies as a form of Big Data. Big Data’s relation to architecture as a practice must also be clarified, as the term “architecture” often refers to the organization of hardware and software systems designed to analyze Big Data. This paper will be referring to both the architecture of Big Data systems as well as the field and profession of architecture.

Another prominent concept that appears in this paper is that of Artificial Intelligence (A.I.). Artificial Intelligence often coincides with conversations about Big Data, as it is an optimal method for analyzing large datasets due to it being able to recognize patterns and relationships on a scale that a human might struggle with. There are various different typologies
and subsets associated with A.I. such as Machine Learning, Predictive Deep Learning, and Natural Language Processing (NPL). Throughout the course of this paper, these different subsets will be mentioned and explored as we overview how current professionals are utilizing A.I. and Big Data to address various challenges present in the field of architecture and design. It is through this analysis that we intend to speculate on how these methodologies could be synthesized to create a process designed for providing inspiration and insight to architects working on any particular project. This process would ideally utilize text and image data to create an output with some sort of spatial or visual quality that an architect could then derive inspiration from. The goal of overviewing these various precedent studies is to help answer the defining question of this research: Through what methods can we extract Architectural Inspiration from Big Data?

**Conducting Evaluation through Big Data Mining**

One of the main reasons for the aforementioned distinction of the term “Architecture” is that much of the research for this paper through scholarly sources would lead us to find articles less focused on architecture in the built environment, and rather on architecture in terms of hardware and software designs within Artificial Intelligence Networks. Although at the surface level these studies may not appear to be related to the topic of inspiration, studies of this nature require a deeper dive due to how they utilize large data sources such as social media platforms. The abstract of Pandey and Purohit enforces the importance of social media data with the following: “Social Media has not only become a platform for mere entertainment and communication but a great source of innovation for public service” (Pandey and Purohit, 2018). Through their research, they proposed a framework for creating a social media data-driven “Intervention Support System (ISS),” which would model trends in the city’s public opinion towards certain campaigns (Figure 1). By viewing the figure, we begin to get an understanding of the inner workflow of data mining that comprises their methodology. They then applied this data collection method to developing an ISS aimed at reducing instances of gender-based violence within the framework of a smart city. This is just one example of how mined data from social media can be utilized in the production of something new, and one could argue that this can be applied to the creative architectural process.
The use of data mining or crawling has also been explored in the context of analyzing urban trends in smart cities. Taking our analysis into the realm of Urban Design only further reinforces the feasibility of incorporating this sort of workflow into the architectural design process, which will be touched upon later in this paper. One such study, Chen et al, was aimed at determining what characteristics define urban areas where innovation is occurring; several of these areas have begun to appear in cities across the United States, so the study oriented its data-gathering efforts into crawling multiple social media sources to determine what these characteristics were to help urban designers plan more successful future spaces (Chen et al, 2016). In terms of the methodology utilized in the study, they specified how they utilized each social media platform with the following: “we utilized geo-located tweets from Twitter, pictures from Flickr, and the density of amenities from Yelp.” In addition to these platforms, Crunchbase (which is an online platform where start-up companies can list themselves, their location, employment data, etc.) was also used to geolocate and visualize various business headquarters and start-up locations on a map of Boston (Figure 2).

From this geospatial baseline, they were able to use Yelp as a source to gauge the number of amenities such as restaurants and grocery stores that were located around or within these various innovation districts. The targeted use of data gathered from multiple social media platforms to assess what elements of these districts were viewed positively by the public exhibits the true potential that a multimodal approach can hold when analyzing streams of Big Data. It was from this approach that Chen et al were able to conclude that the area of Kendall Square was viewed by the public as lacking in public amenities, and subsequently, they were able to propose urban design solutions. The results of this study reveal a critical realization that will only further support the assertions of this paper; the mining and analysis of Big Data are shown to yield data that can influence and drive architecturally designed solutions.
For the results of the previous study to remain a credible example for this argument, we offer another example of the capabilities of Big Data and A.I. being employed to acquire information that will prove insightful in creating and offering design solutions. In an article published in 2020 for the annual European Safety and Reliability Conference, the author writes about how they channeled the interplay between human and machine intelligence to provide architecturally sound suggestions to improve areas of disaster relief and response (Saldaña Ochoa, 2020). This approach used text analysis to scour multiple architectural literary abstracts that were focused on the field of disaster response in order to construct a data set. Humanitarian Relief organizations were also gathered through ReliefWeb and Wikipedia, and these organizations were subsequently merged into a coherent dataset (Figure 3). A neural network algorithm known as Word2Vec was then used to encode the text data from both the literature abstracts and the relief agency mission statements into numerical vectors. These numerical vectors were then fed into an unsupervised machine learning Self Organizing Map (SOM) algorithm (Khonen, 1982). The SOM was then able to create a visualization of the encoded dataset. This visualization can be described as a shifting spectrum of colored cells, each having its own distinct coordinates along with an associated n-dimensional vector otherwise known as a Best Matching Unit (BMU). Each cell was assigned a color based on its weight value (n-dimensional value) and simultaneously displayed a selection of the most commonly occurring keywords from the mission statements and abstracts allocated to the respective cell (Figure 4).
The above dataset was then filtered manually, by a human user, and certain cells were selected from this visualization and extracted to then train a smaller SOM. This step is critical in linking this study to the previous, as they share the common aspect of human intervention and decision-making at a particular stage in the research. The sentiments of these two studies will
carry down to how these workflows can be utilized for the purpose of architectural inspiration and exploration. Though it should be noted that the secondary study takes an approach that could be viewed as daunting to designers who are interested in using a similar workflow but may not have the technical understanding or experience to utilize them effectively. Although these studies were not directly geared toward the topic of providing architectural inspiration, it can be asserted that the workflows that they employ have the potential to be adjusted and honed toward a search aimed at providing a designer with inspiration or insights. This concept will be explored further when we synthesize the methodologies of further studies.

**Analyzing Spatial Social Behavior**

To further develop our literature analysis, we felt it important to subdivide this section into studies that looked at social media as an analytical tool to understand social behavior in a new light. Many of these studies tend to put some emphasis towards the spatiotemporal aspects of social behavior; they explored how social media data could provide insights into how people socialize or act in particular spaces. Social behavior is pivotal to architectural design for its ability to help designers gauge the successes or shortcomings of certain spatial organizations, on both the scale of an urban setting or even an individual building or structure.

One such study uses data collected from Twitter users across Singapore to gauge the diversity level of certain social spaces (Chuang, 2020). The basis of the study relies on the geospatial data associated with where individuals sent tweets along with the determination of their “home location,” which is used in the assessment of a space’s diversity measure; the more people from different neighborhoods visit a location, the more diverse that location is (Figure 5). This geospatial data allowed Chuang to assign certain spaces as “social hubs” based on the measured traffic to that space. This is an insight that would most likely be unseen in a traditional analysis of these hubs that only take into consideration the surrounding population density of an area (Chuang, 2020). Social media data was critical to the identification of these home locations, and by extension, to the categorization and identification of these social spaces existing regardless of the actual designed itinerary or function of the space. Another study takes this utilization of social media data and steps it down to the scale of individual households and apartments, and speaks on the influence that viral fame can have on the very spatial organization of a living arrangement.
Based on our nine-month long ethnography at eight residential areas in Dhaka, Bangladesh, this paper reports how online fame drives local users to produce digital images of their houses mimicking various Western standards, which in turn, brings changes to the organization, aesthetics, and functions of domestic spaces.

(Mim & Ahmed, 2020)

This correlation between online image sharing and the subsequent shifting of apartment layouts by Bangladesh residents proves to be a fascinating insight into how social media has affected the human race spatiotemporally. The study even went to the lengths of interviewing individual families and taking detailed mappings of the apartment floorplans and furniture layout. Several case studies were outlined along with floorplans to visualize the change in circulation that has happened in these apartments due to the clash between cultural practices and social media use (Figure 6). It should be noted that the findings of this study cannot be generalized to any location or population as the results are very dependent on the participants and their living arrangements, however, the methodology of this research is what is significant for its depth and specificity. As was mentioned in the section focused on Big Data analysis, these studies were not oriented toward this research’s focus on the derivation of architectural inspiration, but we can witness the correlation and similarities in methodology that would lend themselves to the use of user data in a SOM analysis, which is prevalent to this paper’s subject.

<table>
<thead>
<tr>
<th>Table 1. Table 1: Homelocation Detection Variables.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Min. tweets sent per user</td>
</tr>
<tr>
<td>2 Min. location sent per user</td>
</tr>
<tr>
<td>3 Min. tweets sent per specific location per user</td>
</tr>
<tr>
<td>4 Min. active hour per specific location per user</td>
</tr>
<tr>
<td>5 Active min. different days per specific location per user</td>
</tr>
<tr>
<td>6 Active min. consecutive days per specific location per user</td>
</tr>
<tr>
<td>7 Removing bots</td>
</tr>
</tbody>
</table>

There are various other examples of social media data being used so that it can aid in the spatiotemporal and social analysis of spaces. One study, in particular, used Twitter data to train an algorithm to begin to understand depressive symptoms related to the COVID-19 pandemic on a spatiotemporal scale in the U.S. (Li et al., 2020). There is even an example of this in a study where social media data was mined in the context of a smart city to craft a digital platform focused on citizen engagement (Sayah, 2019). There are countless other studies being published across various disciplines that are all making use of Social Media and Big Data mining to provide invaluable data for their research. In regards to our topic, the conclusion of the first study overviewed in this section only further enforces the assertion that these workflows are pertinent and relevant within the field of architecture.
In this way, we can provide architects and urban designers with a quick way of pinpointing underperforming areas or well-performing places to investigate for solutions and inform better planning for our future social places in our urban environment. (Chuang, 2020)

There are already a number of different examples across the field of architecture and other design disciplines of this integration, though they are often aimed at being more integrated at the scale of computational design decisions, building analysis, and iterative design techniques (Castro Pena et al., 2021). Before reaching this paper's conclusion we will delve into some of the current uses of A.I. in the field of architecture and how some of the aforementioned workflows and methodologies could be adapted to suit an architect’s search for design insights and inspiration. It should be noted, however, that a major limitation of social media data aggregation is based on user privacy clauses and rights. Certain data is not accessible due to it potentially compromising the privacy of an individual. The proposed framework below then contains the weakness that it is only able to consider data that it has access to, which may limit its ability to produce meaningful results.

**Synthetization of a Methodology**

Given the offered studies that we have analyzed (Table 1), we may be able to look to some already existing examples to formulate a methodology for using A.I. as a catalyst to invoke architectural inspiration. Specifically, we offer an example that looked into integrating a methodology very similar to the study that was aimed at integrating machine and human intelligence in analyzing disaster response. The main exception is that this study was able to take the SOM and integrate the tool into a pedagogical method that was incorporated into an architectural design studio (Saldaña Ochoa & Huang, 2022). This study, once again, utilized text and image data crawled from Twitter to create a SOM to begin the design process for students. Evidence of the work produced during this seminar can be seen in the project boards that students were required to create for the end of the studio (Figure 7). This version of the SOM was also tailored to show the images gathered from Twitter in the cells as well, which was critical in aiding the students when they were asked to create collaged images that depicted the desired atmosphere of their space. Given that this methodology was able to be implemented successfully in architectural pedagogy, the next logical step would be to integrate these
workflows within professional practice. Once the designer or design team has an understanding of their site context and surrounding community, a data crawl using keywords associated with their initial project criteria, and finally utilizing the visualized SOM as a launch point for their architectural design. This is a loose framework suggestion, however, and it could easily be altered to use different algorithms or aggregation techniques, the sole constant would be mining data from a big data source, such as social media.

**Table 1: Data Types and Methodologies of Studies**

<table>
<thead>
<tr>
<th>Reference</th>
<th>Year</th>
<th>Author(s)</th>
<th>Method</th>
<th>Data Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generic Architecture of a Social Media-driven Intervention Support System for Smart Cities</td>
<td>2018</td>
<td>Rahul Pandey and Hernand Punoh</td>
<td>Intervention Support System (ISS) created through data mined from social media and behavioral modeling</td>
<td>Image, Text, and Diversity Data</td>
</tr>
<tr>
<td>Others’ Images: Online Social Media, Architectural Improvisations, and Spatial Marginalization in Bangladesh</td>
<td>2020</td>
<td>Nazir Khan, M. and Syed Inique Ahmed</td>
<td>9 month long on-site ethnography</td>
<td>Image and Text Data</td>
</tr>
<tr>
<td>Sending the Diversity of Social Hubs Through Social Media</td>
<td>2020</td>
<td>I-Ting Chuang</td>
<td>Twitter Data Crawl and Simpson’s Diversity Index algorithm</td>
<td>Text, Image, and Geographic Data</td>
</tr>
<tr>
<td>Social Media as Complementary Tool to Evaluate Cities</td>
<td>2015</td>
<td>Nai Chee Chin, Yukiko Nagahara, Kent Larson</td>
<td>Data Mining: geo-located tweets from Twitter, pictures from Flickr, and the density of amenities from Yelp</td>
<td>Text, Image, and Geographic Data</td>
</tr>
<tr>
<td>Enhancing Disaster Response with Architectural Capabilities by Leveraging Machine and Human Intelligence Interplay</td>
<td>2020</td>
<td>Karla Salavatia Ochoa</td>
<td>Data Crawl: Academic Literature and Humanitarian Organizations Organized by a Self-Organizing Map (SOM) algorithm</td>
<td>Text Data</td>
</tr>
</tbody>
</table>
Aggregation to Inspiration

It should also be briefly noted that at the time of this paper’s writing, there has been an increase in publicly sourced A.I. startups. The emergence of publicly sourced A.I. initiatives such as OpenAI, which is responsible for developing both DALL-E and ChatGPT, may be a positive sign for the future of big data as a source for creativity alongside human input. These programs are in the form of a Neural Network which has the capability of handling both image and text data from billions of data points, and OpenAI as a company has made it their goal to be as open and forthcoming with how they use and develop their tools on their website. These tools are constantly improved upon by the data and queries that they are fed by their users, which is why they are made to be easily accessible to individuals. Only time can determine if and how these tools will develop an interplay within architectural pedagogy and practice.
Conclusion

The main question this research was aiming to resolve was centered around what types of methodologies utilizing A.I. could be used to help aid in providing architectural inspiration. Although many different studies were outlined in this paper to offer possible methodologies, we feel that the visual qualities and breadth of data that Self Organized Mappings are able to show aid themselves to further exploration in the realm of architectural design. This assertion is supported by the elements built into the SOM framework that allow for the input of a human selection to then retrain a more selective SOM based on the selection. The sheer scope of data that can be displayed in this refined visual format opens the door for a variety of images and text data aggregated from social media to be displayed to a designer. It is without question that the strength of this tool is reliant on the user to actually find information that is pertinent to them, which serves as a potential limitation when using these tools. The use of big data solely for the use of inspiration, rather than a physical manifestation of a design, is one proposition for taking advantage of such rich data sources. There are several other plausible routes that mining Big Data can lead to in terms of architectural design, as there are even examples of individual users' data being mined and, through a DLA algorithm, represented in a three-dimensional space (Meekings & Schnabel, 2016). Ultimately, the use of Big Data and A.I. algorithms have shown to be powerful tools that hold the potential to yield new insights to architectural designers, that may even lead to a design’s inspiration.

Acknowledgments

I would like to thank Dr. Karla Saldaña Ochoa for encouraging me to dip my toe into a subject I was very unfamiliar with and undertake this research after having taken her Design Studio in Fall of 2021. Her passion and dedication are greatly appreciated.

References


Chen, N. C., Nagakura, T., & Larson, K. (2016). Social Media as Complementary Tool to Evaluate Cities—Data Mining Innovation Districts in Boston. Herneoa, Aulikki; Toni Österlund and Piia Markkanen (Eds.), Complexity & Simplicity - Proceedings of the 34th ECAADe Conference -


