Proceedings of the 3rd International Conference on Computing Innovation and Applied Physics
DOI: 10.54254/2753-8818/34/20241179

Research on the current status and trend of the application of machine learning in the field of literature and museums—Based on CiteSpace

Jiayuan Chen

Nanjing University of Finance and Economics: No.3, Wenyuan Road, Xianlin University Town, Qixia District, Nanjing city

57314134@qq.com

Abstract. With the development of the world economy and the rapid arrival of the digital era, machine learning technology continues to develop and is widely used in various fields, including the field of literature and museums. The article analyses and researches the relevant literature through citespace visualization software with machine learning research in the field of literature and museums as the main hotspot. It is found that in recent years, the attention to machine learning in the field of literature and museums has shown a significant growth trend, and the machine learning technology is concentrated in the conservation of cultural relics, museum management, cultural heritage protection, heritage identification, etc., which realises the automation, intelligence and personalization of the field of literature and museums and greatly improves the efficiency of work. Finally, this study summarises the current trend of machine learning research in the field of cultural and museums, and puts forward the outlook for the future development direction. By deeply analysing the current situation of machine learning research in the field of literature and exposition, it provides useful references and insights for future researchers, and helps to promote the in-depth development of the intersection between the field of literature and exposition and machine learning, and promotes the innovation and cooperation of related research.

Keywords: Artificial intelligence, machine learning, cultural relics conservation, cultural relics identification, museum management

1. Introduction

In today's age of information explosion, Artificial Intelligence (AI) technology, particularly machine learning, has become a key force driving development in several fields. Machine learning, the core of AI, enables machines to make decisions and predictions without explicit programming by giving computers the ability to learn and mimic human learning. The rapid development of this technology is not only having far-reaching effects in the fields of business, healthcare and education, but is also showing great potential in the cultural and museum fields. The field of culture and museums has traditionally relied on the knowledge and interpretation of artefacts by human experts, but with the intervention of machine learning technologies, the field is undergoing a revolution. From digitally conserving and classifying artefacts to providing customised visitor experiences, the application of machine learning is changing the way people interact with cultural heritage. By sorting out the historical

DOI: 10.54254/2753-8818/34/20241179

data and predicting the future trends, we can promote the innovation and development in the field of literature and museums. The purpose of this study is not only to provide a panoramic view of the application of machine learning in the field of arts and culture, but also aims to provide insights for researchers, practitioners and policy makers to better understand the current state of the field and possible future directions. Through this study, we expect to provide theoretical support and practical guidance for digital transformation and technological innovation in the field of arts and culture.

2. Research methodology and data sources

2.1. Research Methods

A knowledge graph illustrates the intricate relationship among multiple information units or information gatherings, including the structure, interaction, intersection, and evolution of the network; it is a novel technique and area of scientometrics that has gained significant traction in literature searches. Knowledge graphs can be utilized to investigate the development status, the research areas of interest, the frontiers, and the trends and have many advantages, such as being intuitive to use and producing objective results. In addition, through the CiteSpace visualisation tool, keywords were analysed for the screened literature, and intuitive knowledge graphs such as keyword co-occurrence, keyword clustering, and emergent words were generated to identify themes and trends within the field of literature and blogging research, so as to derive the key applications of machine learning in the field of literature and blogging and the direction of future development.

2.2. Data Sources

The data in this paper were obtained from the WOS (Web of Science) database. Web of Science is a comprehensive, multidisciplinary core journal database that covers the top three authoritative citation index databases globally and contains data on hundreds of national and local organizations. To a large extent, the documents indexed by WOS can reflect the application of machine learning to cultural heritage in various countries around the world. Therefore, this paper indexes the literature with the keywords of "machine learning", "cultural heritage", "museum", etc., and a total of 12, 105 documents are retrieved. In order to ensure the timeliness and relevance of the research, a total of 208 documents between 2019 and 2024 were selected as the specific research samples of this paper.

3. Overview of machine learning applications in the field of arts and culture

3.1. Preservation and restoration of cultural heritage

3.1.1. The role of machine learning in heritage identification. The identification of cultural relics can confirm the authenticity of the relics, the era, the maker and other key information, which can help to promote cultural research and education, and enhance the public's knowledge and understanding of cultural heritage, and the introduction of modern analytical instruments combined with machine learning provides a more accurate and efficient method of identification of cultural relics, such as mass spectrometers, spectroscopes and so on, which are able to provide information on the material composition and structure of cultural relics. The multi-dimensional data generated by these instruments serve as the basis for cultural relic identification and are processed by machine learning algorithms to achieve accurate identification of cultural relics.

3.1.2. Computer vision-based restoration technology for cultural relics. As part of cultural heritage, cultural relics carry information about history, art and culture. Through the restoration of cultural relics, these valuable cultural heritage can be protected and inherited so that they can continue to this day and provide rich historical and cultural resources for future generations, while computer vision-based cultural relic restoration technology provides powerful support for cultural relic restoration work through high-precision imaging, such as 3D scanning, high-resolution photography, etc., which is able

DOI: 10.54254/2753-8818/34/20241179

to capture microscopic details of the surface of cultural relics, including cracks, colours and textures, and is able to more accurately detect the damage and absence on the surface of cultural relics, including cracks, missing fragments, etc., which helps restorers to record the current state of cultural relics in detail and provides important diagnostic information for the restoration work.

(OCR) technology supports museums in building digital archives, making it easier to store, manage and share the digitised results of ancient documents, and better protect and pass on cultural heritage. OCR technology can be used to convert handwritten texts or printed documents of ancient books into computer-readable digital texts, which makes it easier to store, retrieve, and share the ancient documents, and provides researchers with wider Access. Image recognition can be used to automatically classify and label images of cultural relics. Through the training model of machine learning, the system can identify different types of cultural relics and automatically generate relevant labels for them to build an orderly digital archive, which helps to manage the large number of cultural relic images in the digital archive.

3.2. Digital management of museums

As they investigate how to use VR/AR technologies to naturally engage with visitors in public settings, museums should also be included in VR/AR learning environments. This will help them fulfill their mandate to impart knowledge of cultural heritage. Using AR and VR technologies, museums can create virtual tours that allow visitors to explore virtual exhibition spaces through smart or specialised devices. This approach provides more flexible pathways while making the tours more interesting. Museums can also create virtual interactive artworks. Visitors can touch, move or sound to create interactive effects and increase engagement. The Song Dynasty Stone Carving Museum in Lu County, Sichuan Province, restores the scene of the Song Dynasty through holographic projection, digital media, AR and VR, and other modern technologies, so that the audience can immerse themselves in the commercial prosperity of the Song Dynasty.

4. CiteSpace visualisation and analysis

4.1. Analysis of core authors

Through the analysis of the co-occurrence map of important authors, it can be found that the main researchers (i.e. authors with high publication volume) in the application of machine learning in the field of literature and blogging are Candela, Gustavo, Liew, Chern Li and so on. Candela, Gustavo published six articles, Liew, Chern Li published four articles, Benford, Steve also published four articles, and other authors basically published 1-2 articles. Culture is like water, infiltrating without sound, connecting the past, present and future of a nation. The preservation and inheritance of culture has always been a topic of great concern to all countries, and many researchers play a pivotal role in this field.

With the help of CiteSpace software, 208 data were analysed by selecting the node type as author's visual graph (Figure 1). Macroscopically, the large font size of the names of authors such as Candela, Gustavo, Liew, Chern Li, etc. in the graph indicates the high frequency of their appearance in this sample of 208 documents. The size of the nodes in the graph represents the number of published papers, and the lines between the nodes indicate the collaborative links between the authors, and the line density is positively correlated with the strength of the collaboration.

DOI: 10.54254/2753-8818/34/20241179

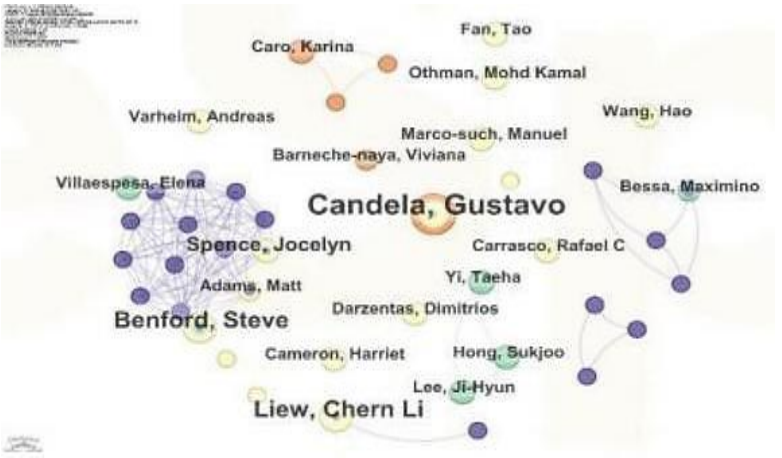

Figure 1. Network map of the core authors

4.2. Analysis of institutional cooperation

In this paper, the network diagram of research institutions is used to show the cooperation between research institutions, and at the same time, the network diagram can also reflect the current situation of research institutions and the close degree of cooperation between institutions and institutions. Among them, the nodes, i.e., the position nodes where the institutions appear, the more nodes, the more papers this institution has published, the connecting line between the nodes represents the connection between the institutions, and the thicker the connecting line, the higher the frequency of them appearing in the same piece of literature. By analysing the WOS literature sample (as in Figure 2), there are 75 nodes with 26 connecting lines in the graph, with a network density of 0.0094. Victoria University Wellington, Universitat d'Alacant, Consiglio Nazionale delle Ricerche (CNR) These three institutions all publish papers with a frequency of 5, and the frequency of publication of the three institutions, namely, the Wuhan University, the University of Nottingham, and the Nanjing University, is 3.

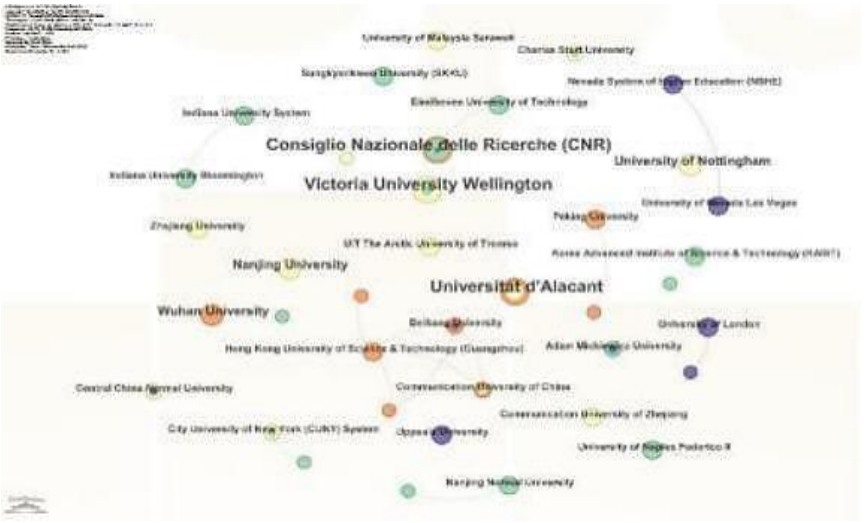

Figure 2. Institutional cooperation network

5. Analysis of Research Hotspots

When analysing research dynamics in a particular academic field, knowledge mapping relies on core metrics such as keywords and the frequency of literature citations. These keywords frequently appear in the titles, abstracts, topics and full texts of research literature, representing the core issues and directions of the field. By quantitatively analysing the frequent occurrences of these terms and establishing a co-

DOI: 10.54254/2753-8818/34/20241179

occurrence relationship map between keywords, the main research trends, development patterns and concerns in the academic field can be effectively revealed. Using this approach, the keyword co-occurrence network constructed in this paper by analysing the 208 WOS literature samples screened can visualise the frequency of occurrence of different keywords, with each node in the network representing a keyword, thus enabling the distribution of high-frequency keywords in the academic field to be clearly identified, and revealing hotspots, evolutionary rules and trend development of research in the field.

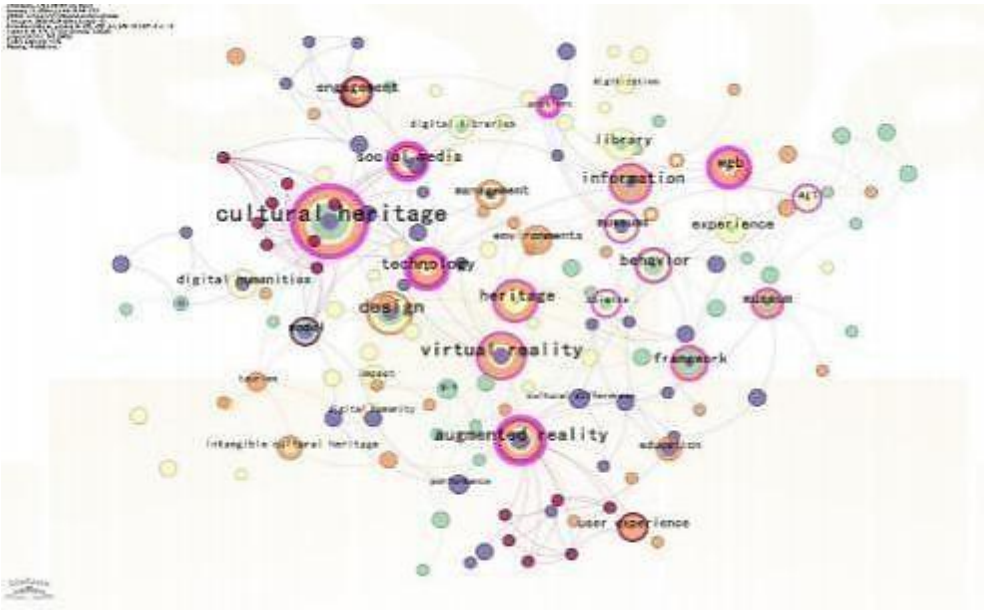

Figure 3. Keyword co-occurrence map

5.1. Keyword Cluster Analysis

A research focus aspect and hotspot can be seen in the keyword co-occurrence timeline map with clusters for the following two reasons: first, publications with relatively strong internal relations are grouped together and thus form a cluster; second, keywords represent and generalize the core idea and research contents of publications to some extent; those with high frequencies are appropriate to track the most recent research hotspots at various research periods .In keyword clustering map, it is generally considered that the Q value is greater than 0.3 means that the clustering structure is obvious, in this map, the Q value is 0.6894 and greater than 0.3, then we can consider that this keyword clustering map clustering structure is obvious, at the same time, the S value is greater than 0.7 means that the clustering is convincing, the S value of this map is 0.9223 is greater than 0.7, so that means that this clustering is convincing. From the clustering map, we can see that there are 9 clusters formed: # digital libraries, # cultural diplomacy, # virtual museum, # social media, # virtual environments, # serious game, # enhancing accessibility, # intangible cultural heritage, # beautiful region, and the clustering modules they form have a certain fit with the co-occurrence frequency.

DOI: 10.54254/2753-8818/34/20241179

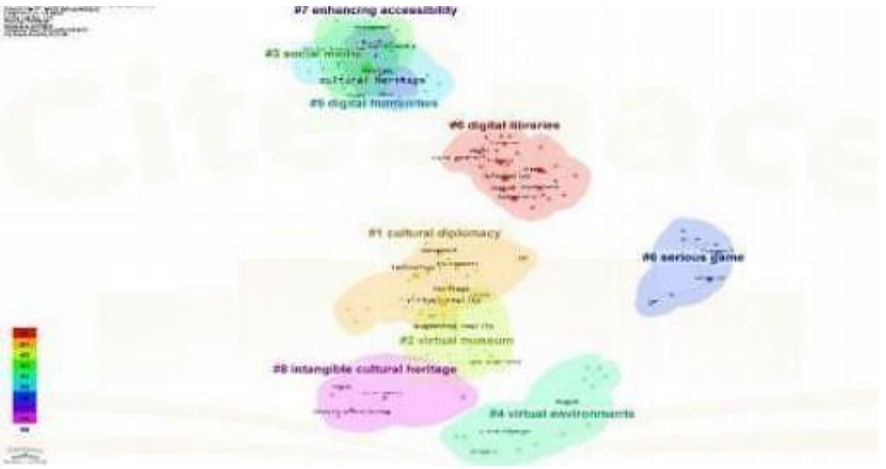

Figure 4. Keyword clustering diagram

5.2. Keyword emergence analysis

Keyword burst mapping is an effective way to show trends in a research topic or keyword over time. Citespace's burst mapping uses gradient colours to show time, with the colours in the mapping changing from light blue (early) to red (mid-term) to dark blue (recent). Figure 5 shows the top 10 keywords with the highest number of citation bursts from 2019-2024. The results show that the top five most explosive keywords are, in order, "Library" (2.03), "Experience" (1.77), "Engagement "(1. 19), "Digital libraries" (1), and "Science" (1), which are the keywords that researchers have paid special attention to, and which are likely to become the most important keywords for the

research on the application of machine learning to the field of literature and blogging. learning's application to the field of arts and culture and blogging is a turning point in research. The top 4 keywords with the longest outbreak are "Engagement" (3 years), "Intangible cultural heritage" (3 years), "Heritage" (3 years), and "Accessibility" (3 years), which implies that researchers have been focusing on these keywords for machine learning applications to the field of literature and blogging over the time period in which the literature sample was taken.

Top 10 Keywords with the Strongest Citation Bursts

Keywords	Year	Strength	Begin	End	2020 - 2024
model	2020	0.76	2020	2021	
artificial intelligence	2020	0.48	2020	2021	
library	2021	2.03	2021	2022	
experience	2021	1.77	2021	2022	
digital libraries	2021	1	2021	2022	
science	2021	1	2021	2022	
engagement	2022	1.19	2022	2024	
intangible cultural heritage	2022	0.79	2022	2024	
heritage	2021	0.77	2022	2024	
accessibility	2022	0.39	2022	2024	

Figure 5. Keyword emergent word spectrue

DOI: 10.54254/2753-8818/34/20241179

6. Conclusion

Through a Citespace-based study, we provide an in-depth analysis of the current status and trends of machine learning applications in the field of literature and museums. Firstly, we observe that the application of machine learning in the field of literature and museums shows a rapid growth. Through the visual analysis of Citespace, we clearly see the dense area of related research, which reflects the continuous attention and investment of academics in applying machine learning in the field of literature and museum. This demonstrates that machine learning research in the field of arts and crafts has become a hotspot of great interest. Second, we note the trend of multi- domain cross fertilisation.Citespace's analysis reveals that machine learning is interacting closely with other subdomains in the field of arts and culture and blogging through increasingly frequent cross collaboration, such as in artificial intelligence. This multi-domain convergence has fuelled the emergence of new research directions and methods, providing new impetus for innovation in the field of arts and culture. In the course of our research, we are also concerned with a number of challenges and issues.

References

[1] Building of Digital Archives of Kyrgyz Museums: Preserving the Cultural Resources of Kyrgyzstan. (2023). Hanhee Hahm, Semina Oh, Minhee Choi, Soon Cheol Park.https://doi.org/10.18178/wcse.2023.06.044

[2] Calero-Sanz, J., Orea-Giner, A., Villacé-Molinero, T., Muñoz-Mazón, A., & Moraleda, L. F. (2022). Predicting A New Hotel Rating System by Analysing UGC Content from Tripadvisor: Machine Learning Application to Analyse

[3] Service Robots Influence. Procedia Computer Science, 200, 1078– 1083. https://doi.org/10.1016/j.procs.2022.01.307

[4] Dağ, K., Çavuşoğlu, S., & Durmaz, Y. (2023). The effect of immersive experience, user engagement and perceived authenticity on place satisfaction in the context of augmented reality. Library Hi Tech. https://doi.org/10.1108/lht- 10-2022-0498

[5] Geng, Y., & Maimaituerxun, M. (2022). Research Progress of Green Marketing in Sustainable Consumption based on CiteSpace Analysis. SAGE Open, 12(3), 215824402211198. https://doi.org/10.1177/21582440221119835

[6] Ronco, F. (2021). Digital technologies applied to the accessible management of museums. The first experiments carried out at the Museum of Oriental Art of Turin. Revista De Expresión GráficaEn La Edificación, 14, 49. https://doi.org/10.4995/ege.2021.15661

[7] Scavarelli, A., Arya, A., & Teather, R. J. (2020). Virtual reality and augmented reality in social learning spaces: a literature review. Virtual Reality, 25(1), 257-277. https://doi.org/10.1007/s10055-020-00444-8

[8] Song, M., Zhang, J., Liu, X., Zhang, L., Hao, X., & Li, M. (2023). Developments and Trends in Energy Poverty Research—Literature Visualization Analysis based on CiteSpace. Sustainability, 15(3), 2576.