Metadata-enriched Discovery Of Aspect Similarities Between Cultural Objects

Yorghos Voutos
Ionian University
c16vout@ionio.gr

Georgios Drakopoulos
Ionian University
c16drak@ionio.gr

Phivos Mylonas
Ionian University
fmylonas@ionio.gr

***Abstract*—Evaluating in the digital realm the similarity between two cultural objects relies on a number of factors. The latter can be subdivided to three categories, each describing a different set of properties. Specifically, the first such category pertains to generic attributes which usually are domain oblivious. For instance, for literature works standard text similarity metrics based on attributes such as n-gram or word frequencies. The second category consists of attributes deriving from domain-specific aspects. In the same example, a literature work may be described by the style, the author, and the year it was written. Finally, the third category relies on metadata generated by humans who have previously accessed the specific cultural object, perhaps drawn from a pre-specified list to avoid cluttering by redundant, erroneous, or invalid descriptions. Despite this restriction, the latter allows for an enhanced and more accurate description of the cultural object as a human can associate quicker and more efficiently a certain set of descriptions to it. Here a metric is proposed based on all three categories and it is compared against baseline ones with encouraging results.**

***Index Terms*—digital culture, cultural content, cultural heritage management, user annotation, metadata, cultural clustering, attribute engineering, clustering quality, distance metrics**

1. Introduction

Digital culture and its associated content creation and management in general is undoubtedly one of the mainstays of the modern digital era. In fact, it has been recently estimated by multiple independent authors like those of [1] [2] [3] that digital cultural content creation rate has already surpassed that of ordinary cultural content creation, whereas the demand for such content has sharply risen the past years [4]. A number of explanations have been proposed to the latter including the additional leisure time of the average citizens of developed countries [5] to the identity building efforts of the generation Z age cohort [6] [7]. Regardless of the major drivers behind it, the spike in the demand for digital cultural content results in a need for sophisticated cultural analytics such as recommendation systems, demand predictors, and adaptive delivery content algorithms to name just a few. In this context cultural object similarity metrics play a central role. The design of such metrics is by no means a trivial task as there is a plethora of attributes to factor in. For instance, for a digitized painting a minimal set of prominent features would include the creator, era, style, topic, main colors, original material, resolution, and perhaps a brief description of any restoration efforts.

One significant element of the mindset of the digital culture ecosystem is the growing and active participation of digital consumers in various capacities. Instead of restraining themselves to commenting on places like dedicated *fora*, cultural portals, and social media, digial consumers now may well do more than that. Individual independent producers on a daily basis create videos to be posted in platforms such as YouTube and Vimeo, upload podcasts to Spotify, or post digital art on Tumblr. Another side of this activity, less complicated but nonetheless important, is the manual annotation of cultural objects, typically in the form of hashtags, annotations, or short descriptions. Such human generated metadata often contain valuable semantic information which cannot in general be obtained by ordinary classification or artificial intelligence (AI) algorithms.

The primary research objective of this conference paper is twofold. First, a model for placing cultural object attributes in three distinct groups is proposed. These groups are generic object features, domain specific attributes, and annotations from digital consumers. Second, a similarity metric is proposed based on this categorization. The latter is generic enough to cover most cultural object class supporting metrics and also customizable through a small number of weights defined in an intuitive friendly way. In turn, as a concrete example this metric is applied to clustering objects from a dataset obtained from a specific cultural game. Results from standard clustering quality evaluation techniques indicate the proposed methodology is promising.

The remaining of this conference paper is structured as follows. Section 2 reviews the recent scientific literature regarding digital culture management, digital culture portals, and cultural object manual annotation and metadata management. Section 3 describe the proposed methodology and section 4 presents the dataset and the experimental results. Section 5 concludes this work by recapitulating the main findings of this work and outlines possible future research directions. Technical acronyms are explained the first time they are encountered in the text. Function parameters are placed after a semicolon following its arguments. The terms *user* and *digital consumer* will be used interchangeably.

978-1-7281-5919-5/20/$31.00 ©2020 IEEE

Matrices are represented by boldface capital letters and vectors by boldface small ones. Finally table 1 summarizes the notation of this work.

TABLE 1. NOTATION OF THIS CONFERENCE PAPER.

Symbol	Meaning
$\triangleq$	Definition or equality by definition
$\{s_1, \ldots, s_n\}$	Set containing elements $s_1, \ldots, s_n$
$(t_1, \ldots, t_m)$	Tuple with elements $t_1, \ldots, t_m$
$\lvert S \rvert$	Set cardinality
$\mathrm{diag}\,[d_1, \ldots, d_n]$	Diagonal matrix with elements $d_1, \ldots, d_n$
$\mathrm{E}\,[X]$	Mean of random variable X
$\mathrm{Var}\,[X]$	Variance of random variable X

2. Previous Work

Digital culture is an emerging domain seeking to agument the existing social cultural dissemination and interaction routes with digital means [8] [9]. Currently the digital culture ecosystem is the heart of a specialized economy whose main traits are openess, decentralization, sharing of resources, emphasis on innovation and creativity, digital markets and auctions, and a bottom up operational attitude [10]. Aspects of digital culture from a business perspective are given in [11], while the particular role and perspectives of traditional cultural industries in general to the new digital economy are explored in [12]. Cultural content description is vital for successful communication [13] [14] and content delivery [15]. To this end established open data formats such as Javascript Object Notation (JSON) for cultural metadata can be used [16]. The role of the *digital natives*, a composite group consisting mostly of members of generations Y and Z, and how they perceive digital culture are the focus of [17]. The participation of local communities to cultural heritage preservation through digitization is examined in [18].

Culture oriented software, including cultural heritage software and cultural games, is currently becoming popular for a number of reasons as explained among others in [19] and in [20]. Cross-cultural design considerations for mobile apps are given in [21]. Along a similar line of reasoning, cross-cultural principles for game design are the focus of [22]. Additionally, cultural games are now considered a mainstream branch of gaming [23] [24]. Ontologies for personalization [25] provide a formal methodology for making these games more appealing and personalized. Also, cultural analytics based on tenets from cognitive sciences and behavioral economics typically are the backbone of these games [26]. It should be noted that cultural games and games in general are now starting to be considered as part of the global human heritage [27].

Clustering is one of the fundamental data mining techniques [28]. Strategies for clustering range for objects like graphs [29] and strings [30] to numerical vectors [31]. Over time numerous schemas have been proposed, mainly for the hierarchical [32] or the agglomeratvie clustering paradigm [33]. In fuzzy clustering objects way well belong to more than one cluster [34]. Moreover, ontologies for visual management [35] may be the key for human assisted clustering since humans rely heavily on visual cues. Recent applications include tensor clustering [36], ranking Twitter accounts based on a social tensor containing functional and structural attributes [37], and opinion mining based on product aspects [38].

3. Proposed Methodology

3.1. The Model

The proposed methodology relies on combining attributes from three distinct factors, each of which captures different aspects of a given cultural object, in order to derive a simialrity metric for cultural objects. Specifically:

- The *object attributes* are domain oblivious and focus on the object itself as an entity.
- The *domain attributes* pertain to these aspects which define the cultural dimension of the object.
- The *annotations* come from the digital consumers accessing and using the cultural objects.

Since the numerical range of the features from the above factors may be different or the contribution of each factor may vary, a normalization process is necessary in order to combine the factors. To this end, a verification process, perhaps by a human expert or by a dedicated algorithm, is in order to ensure that normalization is meaningful according to the properties of the underlying domain.

Figure 1 depicts the basic concepts of the proposed model and their interactions.

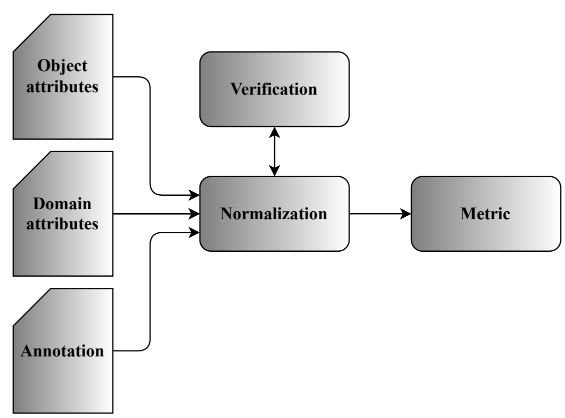

Figure 1. Factors of the proposed metric construction model.

Since cultural items may well have more than one modality, for instance a video showing folk dances or a poet reading his/her own poems, a text about art or a literary review, or a painting showing a historical monument, the proposed model can be naturally extended to each modality. In this case, each such modality may be separately used or an averaging scheme may be utilized.

3.2. Object Attributes

Despite the particular nature of cultural objects, they still reamain parts of the digital domain and as such they have the respective generic properties. For instance, digitized paintings or photographs have a specific color range distribution in the red-green-black (RGB) or in the cyan-magenta-yellow-black (CMYK) systems or a gradient coding allowing region identification. Similarly, a document may well be described in terms of the n-grams, word frequency and word length distributions, or the total number of characters. Object attributes can be used in order to probabilistically determine work authorship or era.

Table 2 shows some of the most common generic attributes for the popular forms of cultural objects.

TABLE 2. INDICATIVE OBJECT ATTRIBUTES

Object	Attribute
Music	Spectrogram, octaves and scales used
Text	n-gram, word frequency and word length distributions
Pictures	Color distribution, compression scheme
Video	Resolution, color distribution, compression scheme

3.3. Domain Attributes

Any cultural object ultimately represents certain values like truth or beauty and expresses a past or a contemporary artistic movement. Moreover, it certainly has a creator, even a collective or an anonymous one as is the case for folk tales and songs, as well as a specific historical era of creation, which often provides context for the object. Variations, typically linguistic or artistic, also play a significant role in assessing both the effect of a cultural object to the collective culture and the spread of a given work in space and time.

Table 3 contains certain indicative features for the most common types of cultural objects.

TABLE 3. INDICATIVE DOMAIN ATTRIBUTES

Object	Attribute
Music	Composer, number and type of organs, era
Text	Language, author(s), movement, topic
Pictures	Photographer, era, processing type, camera
Video	Creator, era, topic, movement, processing

3.4. Cultural Object Annotation

Since in actual cultural portals the various available objects are eventually used by cultural consumers, either through physical or digital interaction, it makes perfect sense to get feedback from the digital consumer base in order not only to improve the overall user experience (UX), which is a very important topic on its own right, but also to gain insight into how the consumer base sees these objects and into any latent connections between them. Examples of the latter case include object similarity input from digital consumers of diverse ethic, educational, or cultural backgrounds.

Similarities in annotations often indicate non-apparent connections between objects which typically cannot be algorithmically determined such as emotional response or aesthetic significance.

Policy alternatives for the annotations include:

- Allow users to pick annotations only from a fixed list. This option eliminates ambiguity and portal abuse from potentially malicious digital consumers, but also it is the most restrictive. Also, the list should be frequently updated based on user feedback.
- Allow users to enter their own annotations for every cultural object. In this case natural language processing (NLP) and ontology-based techniques should be employed in order to address issues such as duplicate entries or annotations with unclear meaning.
- An intermediate solution would be to let digital consumers write their own annotations indicating at least one category an annotation belongs to. This reduces semantic-related issues while providing users a significant amount of freedom. This choice was implemented in the game which generated the dataset analyzed in the experiments.

3.5. Annotation-Assisted Clustering

Given the preceding analysis, the proposed similarity metric $d(s_1, s_2)$ between cultural objects s_1 and s_2 is mathematically formulated as the linear combination of (1):

$$\begin{aligned} d(s_1, s_2) \triangleq\ & w_o g_o \left(h_o(s_1, s_2); 0, \sigma_g^2\right) \\ & + w_d g_d \left(h_d(s_1, s_2); 0, \sigma_d^2\right) \\ & + w_a g_a \left(h_a(s_1, s_2); 0, \sigma_a^2\right) \end{aligned} \tag{1}$$

The one dimensional Gaussian kernel $g\left(x; \mu_0, \sigma_0^2\right)$ is the basic buidling block of $d(\cdot, \cdot)$ and it is defined as in (2):

$$g\left(x; \mu_0, \sigma_0^2\right) \triangleq \frac{1}{\sigma_0\sqrt{2\pi}} \exp\left(-\frac{(x-\mu_0)^2}{2\sigma_0^2}\right) \tag{2}$$

The Gaussian kernel of equation (2) has been chosen for the following reasons:

- The Gaussian distribution has the maximum differential entropy among all distributions with the same variance [39], meaning that the Gaussian distribution can explain more probabilistic scenarios.
- Any linear combination of independent Gaussian distributions variables is itself a Gaussian distribution [40]. This allows the easy determination of the parameters of the resulting distribution as well as its study in general.
- It is continuous and smooth as such it does not have any jumps or other points with singular behavior [41]. Therefore distance computation can take place throughout the domain of $d(\cdot, \cdot)$.
- Its decay rate, which is low only locally in the neighborhood of up to σ_0^2 but sharp after than point [41], guarantees that only objects which are really

close are assigned a high similarity value. The ratio of two zero mean Gaussian kernels drops below a given threshold γ_0 when:

$$\frac{g\left(x_0; 0, \sigma_0^2\right)}{g(x_1; 0, \sigma_0^2)} = \gamma_0 \Leftrightarrow x_1 = \pm\sqrt{x_0^2 + 2\sigma_0^2 \ln \gamma_0} \tag{3}$$

The weights w_h, w_d, and w_a of equation (1) are non-negative, sum up to one, determine the importance of each individual factor. Assuming $d\left(\cdot, \cdot\right)$ takes into consideration n_o generic object atrributes, n_d domain attributes, and n_a user annotations, there are many ways to assess the individual contribution including the following ones:

- All three weights can have the same value to indicate that each factor contributes equally to the total object similarity. In this case:

$$w_o = w_d = w_a = \frac{1}{3} \tag{4}$$

- Each weight is proportional to the number of attributes beloning to the corresponding factor to the total number of features examined. Therefore:

$$w_o \triangleq \frac{n_o}{n_o + n_d + n_a} \tag{5}$$

  The remaining two weights are similarly defined.
- Alternatively, each weight can be the softmax score defined as:

$$w_o \triangleq \frac{\exp\left(n_o\right)}{\exp\left(n_o\right) + \exp\left(n_d\right) + \exp\left(n_a\right)} \tag{6}$$

  The remaining two weights are similarly defined.

Notice that the mean values of all three Gaussian kernels of equation (1) are zero. This was done on purpose as a non-zero value would introduce bias violating the condition that each of these kernels should take its respective maximum value if and only if s_1 and s_2 have the same attributes. On the other hand, the variances σ_o^2, σ_d^2, and σ_a^2 determine the locality of comparisons in the respective factors.

Finally, the distance metrics $h_o\left(\cdot, \cdot\right)$, $h_d\left(\cdot, \cdot\right)$, and $h_a\left(\cdot, \cdot\right)$ determine the distance between the attributes. Their selection is free as long as the following conditions are met:

- The metric returns a single numerical value.
- The range is $[0, +\infty)$, with 0 reserved only for when the attributes of s_1 and s_2 are the same.

Table 4 has common choices for various attribute types.

TABLE 4. METRIC DISTANCE OPTIONS

Attribute	Metrics
Numerical	ℓ_1 and ℓ_2 norms
Alphanumeric	Inverse Tanimoto, Leacock-Chodorow, Levenshtein

Algorithm 1 shows how proposed metric in the context of k-means, the clustering scheme used in the experiments. Please see section 4 for an analysis of this algorithm.

Algorithm 1 Proposed cultural object clustering in k-means

Require: Weights and distance metrics for equation (1)
Require: Number of clusters q_0 and criterion τ_0
Ensure: The cultural objects $\{s_k\}_{k=1}^{n}$ are clustered

for all clusters $\{C_j\}_{j=1}^{q_0}$ **do**
    pick a random s_j **and** assign centroid $c_j \leftarrow s_j$
end for
repeat
    for all objects $\{s\}_{k=1}^{n}$ **do**
        for all centroids $\{c_j\}_{j=1}^{q_0}$ **do**
            compute $d\left(s_k, c_j\right)$ as in equation (1)
        end for
        assign s_k to cluster with maximum similarity
    end for
    if there are empty clusters **then**
        set q_0 as the new mumber of clusters
    end if
    for all clusters $\{C_j\}_{j=1}^{q_0}$ **do**
        for all $s_k \in C_j$ **do**
            find the average distance $\bar{d}_k$ as in (7)
            pick as c_j the object with the least $\bar{d}_k$
        end for
    end for
until τ_0 is **true**
return clusters $\{C_j\}_{j=1}^{q_0}$

4. Results

The dataset used in the experiments contains n texts from the cultural portal of project . For each such text p_o features were collected, namely word frequency and word length distribution, total number of words, and total number of sentences. Additionally p_d attributes were also recorded, namely author, era, artistic movement,and location, while up to p_a annotations per document were collected.

The experimental setup is shown in table 5. The proposed simialrity metric was used in the k-means scheme shown in algorithm 1. The latter was selected because of its simplicity and flexibility. Notice that in contrast to the standard version the centroids $\{c_j\}_{j=1}^{q_0}$ during each iteration are computed by picking the object with the minimum average distance from each other object in the cluster. This happens because distances between objects can be computed through $d\left(\cdot, \cdot\right)$ but there is no way to compute the average of the annotations. On the other hand, the average distance $\bar{d}_k$ of an object $s_k \in C_j$ from all other objects assigned to the same cluster can be defined as in equation (7):

$$\bar{d}_k \triangleq \frac{1}{|C_j| - 1} \sum_{s_i \in C_j} d\left(s_i, s_k\right) \qquad s_i \neq s_k \tag{7}$$

In equation (7) $|C_j|$ stands for the number of the objects in the cluster. An advantage of this strategy is that centroids are by construction valid objects and therefore can be analyzed once obtained for further insight. Since the centroids are computed in a different way than standard k-means, the termination condition τ_0 for algorithm 1 was that clusters would remain the same for two consecutive iterations.

TABLE 5. EXPERIMENTAL SETUP

Parameter	Value
Number of cultural ojects n	2039
Type of cultural objects	Documents
Object attributes p_o	4
Domain attributes p_d	4
User annotations p_a	5
Cluster initialization policy	Random
Object attribute distance $h_a(\cdot,\cdot)$	Weighted ℓ_2 norm
Domain attribute distance $h_a(\cdot,\cdot)$	Weighted ℓ_2 norm
User annotation distance $h_a(\cdot,\cdot)$	Average Levenshtein
Variances σ_o^2 / σ_d^2 / σ_a^2	p_o / p_d / p_a
Weight policy W	**U**niform/**S**oftmax
Consider annotations A	**Y**es/**N**o
Number of clusters q_0	$\lceil n \rceil$
Number of runs R_0	1000

For the object and attributes which in this particular case have only numerical attributes or logical which can be translates to numerical ones the weighted ℓ_2 norm will be used. In the following $h_d(\cdot,\cdot)$ will be used, but the same procedure holds for $h_o(\cdot,\cdot)$. Let object s_k be represented by the attribute column vector $\mathbf{f}_{k,d}$ of length p as in (8):

$$\mathbf{f}_{k,d} \triangleq \begin{bmatrix} f_{1,d} & f_{2,d} & \dots & f_{p,d} \end{bmatrix}^T \tag{8}$$

Let the weight matrix $\mathbf{M}_d$ be the diagonal matrix of (9):

$$\mathbf{M}_d \triangleq \operatorname{diag}\left[\frac{1}{F_{1,d}^2}, \dots, \frac{1}{F_{p,d}^2}\right] \tag{9}$$

In (9) $F_{k,d}$ is the maximum value of the k-th attribute. The distance $h_d(s_1, s_2)$ between s_1 and s_2 is given then by (10):

$$h_d(s_1, s_2) \triangleq \sqrt{(\mathbf{f}_{1,d} - \mathbf{f}_{2,d})^T \mathbf{M}_d (\mathbf{f}_{1,d} - \mathbf{f}_{2,d})} \tag{10}$$

Since annotations are strings $h_a(\cdot,\cdot)$ has been chosen to be the average Levenshtein distance determined as:

- For each annotation l_i from s_1 and l_j from s_2 the pairwise Levenshtein distance is computed.
- Each l_i is matched to the closest l_j. Any remaining annotations are matched to the empty string.
- The average distance of all these pairwise matches is the value of $h_a(s_1, s_2)$.

The values for variances σ_o^2, σ_d^2, σ_a^2 were selected to be equal to the number of the respective attributes.

Each clustering configuration can be represented as a tuple with the format of equation (11). The parameters are explained in table 5 and actual configurations in table 6. Additionally, the latter table contains the experiment results.

$$(W, A, q_0) \tag{11}$$

Since k-means is probabilistic, each clustering configuration was executed R_0 times and the number of iterations and the average cluster distance were recorded. Then the respective mean values $\mathrm{E}[I]$ and $\mathrm{E}[J]$ as well as the respective variances $\mathrm{Var}[I]$ and $\mathrm{Var}[J]$ were computed.

TABLE 6. CLUSTERING CONFIGURATIONS AND RESULTS

Configuration	$\mathrm{E}[I]$ / $\mathrm{Var}[I]$	$\mathrm{E}[J]$ / $\mathrm{Var}[J]$
(U, N, q_0)	53.6667 / 18.4667	6.5588 / 11.1193
(S, N, q_0)	58.3334 / 17.5113	6.3577 / 12.4574
(U, Y, q_0)	33.1666 / 18.6618	4.1612 / 11.5535
(S, Y, q_0)	34.5000 / 18.1224	3.8999 / 11.6674

The distance $D_{j,j'}$ between clusters C_j and $C_{j'}$ is the mean distance between any pair $s_k \in C_j$ and $s_{k'} \in C_{j'}$:

$$D_{j,j'} \triangleq \frac{1}{|C_j||C_{j'}|} \sum_{s_k \in C_j} \sum_{s_{k'} \in C_{j'}} d(s_k, s_{k'}) \tag{12}$$

Averaging $D_{j,j'}$ over all distinct cluster pairs yields J.

From the entries of table 6 the following can be said:

- The annotations reduced considerably the number of iterations, meaning that clustering is more efficient.
- Moreover, they increased the average cluster distance, meaning that the clusters are more compact and at the same time more discernible.

5. Future Work

This conference paper focuses on the development of a metric for determining similarity between cultural objects for cultural games or portals. The latter depends on three general factors capturing different aspects of these objects. These factors are object attributes, domain features, and user annotations. Object attributes are domain oblivious and pertain to object itself, while the domain features capture its cultural aspects. The annotations come directly from the users and they frequently useful in discovering non-trivial and non-obvious relationships between cultural objects. In this conference paper the proposed metric has been applied in a version of k-means to cluster a collection of documents with and without annotations. The former choice resulted in less iterations and also in an improved clustering quality.

Concerning future research directions, more factors can be added to the proposed model. Moreover, clustering performance with larger and more diverse datasets should be explored. Also, robust clustering techniques should be developed in order to reduce the number of iterations required.

Acknowledgment

This research has been co-financed by the European Union and Greek national funds through the Competitiveness, Entrepreneur- ship and Innovation Operational Programme, under the Call "Research – Create – Innovate", project title: "Development of technologies and methods for cultural inventory data interoperability", project code: T1EDK-01728, MIS code: 5030954.

References

[1] E. Not and D. Petrelli, "Empowering cultural heritage professionals with tools for authoring and deploying personalised visitor experiences," *User Modeling and User-Adapted Interaction*, vol. 29, no. 1, pp. 67–120, 2019.

[2] B. Matheson and E. J. Petersen, "Engaging US students in culturally aware content creation and interactive technology design through service learning," *IEEE Transactions on Professional Communication*, vol. 63, no. 2, pp. 188–200, 2020.

[3] J. D. Snowball, "Cultural value," in *Handbook of Cultural Economics*, 3rd ed. Edward Elgar Publishing, 2020.

[4] J. McKenzie and S. Y. Shin, "Demand," in *Handbook of Cultural Economics*, 3rd ed. Edward Elgar Publishing, 2020.

[5] D. Sivevska, "Leisure time in contemporary society," *International Journal Knowledge*, vol. 13, pp. 437–442, 2016.

[6] C. Giunta, "Digital marketing platform tools, generation Z, and cultural considerations," *Journal of Marketing Development & Competitiveness*, vol. 14, no. 2, 2020.

[7] A. Turner, "Generation Z: Technology and social interest," *The journal of individual Psychology*, vol. 71, no. 2, pp. 103–113, 2015.

[8] V. Miller, *Understanding digital culture*. SAGE Publications Ltd., 2020.

[9] B. L. Dey, D. Yen, and L. Samuel, "Digital consumer culture and digital acculturation," *International Journal of Information Management*, vol. 51, 2020.

[10] M. Banks, "Creative economy, degrowth and aesthetic limitation," in *Cultural Industries and the Environmental Crisis*. Springer, 2020, pp. 11–23.

[11] N. R. Moşteanu, A. Faccia, and L. P. L. Cavaliere, "Digitalization and green economy-changes of business perspectives," in *International Conference on Cloud and Big Data Computing*, 2020, pp. 108–112.

[12] C. Peukert, "The next wave of digital technological change and the cultural industries," *Journal of Cultural Economics*, vol. 43, no. 2, pp. 189–210, 2019.

[13] X. Xu, "Cultural communication in double-layer coupling social network based on association rules in big data," *Personal and Ubiquitous Computing*, vol. 24, no. 1, pp. 57–74, 2020.

[14] A. Hillary, "Neurodiversity and cross-cultural communication," *Neurodiversity Studies: A New Critical Paradigm*, 2020.

[15] M. Casillo, F. Clarizia, G. D'Aniello, M. De Santo, M. Lombardi, and D. Santaniello, "CHAT-Bot: A cultural heritage aware teller-bot for supporting touristic experiences," *Pattern Recognition Letters*, vol. 131, pp. 234–243, 2020.

[16] G. Drakopoulos, E. Spyrou, Y. Voutos, and P. Mylonas, "A semantically annotated JSON metadata structure for open linked cultural data in Neo4j," in *PCI*. ACM, 2019, pp. 81–88.

[17] T. Tran *et al.*, "How digital natives learn and thrive in the digital age: Evidence from an emerging economy," *Sustainability*, vol. 12, no. 9, 2020.

[18] J. Li, S. Krishnamurthy, A. P. Roders, and P. van Wesemael, "Community participation in cultural heritage management: A systematic literature review comparing Chinese and international practices," *Cities*, vol. 96, 2020.

[19] J. Dugdale, M. T. Moghaddam, and H. Muccini, "Human behaviour centered design: developing a software system for cultural heritage," in *International Conference on Software Engineering: Software Engineering in Society*, 2020, pp. 85–94.

[20] S. Bampatzia, I. Bourlakos, A. Antoniou, C. Vassilakis, G. Lepouras, and M. Wallace, "Serious games: Valuable tools for cultural heritage," in *International Conference on Games and Learning Alliance*. Springer, 2016, pp. 331–341.

[21] N. Sun, D. Frey, R. Jin, H. Huang, Z. Chen, and P.-L. P. Rau, "A cross-cultural study of user experience of video on demand on mobile devices," in *International Conference on Cross-Cultural Design*. Springer, 2013, pp. 468–474.

[22] K. Koban and N. D. Bowman, "Further validation and cross-cultural replication of the video game demand scale," *Journal of Media Psychology*, 2020.

[23] L. Ye, R. Wang, and J. Zhao, "Enhancing learning performance and motivation of cultural heritage using serious games," *Journal of Educational Computing Research*, 2020.

[24] A. Y. Tikhonova, D. Makarov, and A. Maltseva, "Active ethnic games as regional cultural heritage promotion tool," *Theory and Practice of Physical Culture*, no. 2, 2019.

[25] D. Vallet, M. Fernandez, P. Castells, P. Mylonas, and Y. Avrithis, "A contextual personalization approach based on ontological knowledge," *Contexts and Ontologies: Theory, Practice and Applications*, p. 35, 2006.

[26] G. Drakopoulos, I. Giannoukou, P. Mylonas, and S. Sioutas, "The converging triangle of cultural content, cognitive science, and behavioral economics," in *MHDW*, ser. IFIP Advances in information and communication technology, vol. 585. Springer, 2020, pp. 200–212.

[27] L. Eklund, B. Sjöblom, and P. Prax, "Lost in translation: Video games becoming cultural heritage?" *Cultural Sociology*, vol. 13, no. 4, pp. 444–460, 2019.

[28] A. Tolba and Z. Al-Makhadmeh, "An improved density-based single sliding clustering algorithm for large datasets in the cultural information system," *Personal and Ubiquitous Computing*, vol. 24, no. 1, pp. 33–44, 2020.

[29] Z. Wang, Z. Li, R. Wang, F. Nie, and X. Li, "Large graph clustering with simultaneous spectral embedding and discretization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[30] T. Jo, "Semantic string operation for specializing AHC algorithm for text clustering," *Annals of Mathematics and Artificial Intelligence*, pp. 1–18, 2020.

[31] P. D. Pantula, S. S. Miriyala, and K. Mitra, "An evolutionary neuro-fuzzy C-means clustering technique," *Engineering Applications of Artificial Intelligence*, vol. 89, 2020.

[32] P. Mylonas, M. Wallace, and S. Kollias, "Using k-nearest neighbor and feature selection as an improvement to hierarchical clustering," in *SETN*. Springer, 2004, pp. 191–200.

[33] N. Ding and F. Farokhi, "Developing non-stochastic privacy-preserving policies using agglomerative clustering," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3911–3923, 2020.

[34] G. Drakopoulos, P. Gourgaris, A. Kanavos, and C. Makris, "A fuzzy graph framework for initializing k-means," *IJAIT*, vol. 25, no. 6, pp. 1 650 031:1–1 650 031:21, 2016.

[35] S. Dasiopoulou, C. Saathoff, P. Mylonas, Y. Avrithis, Y. Kompatsiaris, S. Staab, and M. G. Strinztis, "Introducing context and reasoning in visual content analysis: An ontology-based framework," in *Semantic Multimedia and Ontologies*. Springer, 2008, pp. 99–122.

[36] M. Ashraphijuo and X. Wang, "Union of low-rank tensor spaces: Clustering and completion," *JMLR*, vol. 21, no. 69, pp. 1–36, 2020.

[37] G. Drakopoulos, "Tensor fusion of social structural and functional analytics over Neo4j," in *IISA*. IEEE, 2016.

[38] W. Kessler, R. Klinger, and J. Kuhn, "Towards opinion mining from reviews for the prediction of product rankings," in *Workshop on Computational Approaches to Subjectivity, Sentiment, and Social Media Analysis*, 2015, pp. 51–57.

[39] K. Nigam, J. Lafferty, and A. McCallum, "Using maximum entropy for text classification," in *IJCAI*, vol. 1, no. 1, 1999, pp. 61–67.

[40] X. Jia and B. Zhao, "Algorithm design of combined Gaussian pulse," in *ICASSP*. Springer, 2019, pp. 1262–1266.

[41] R. Feghhi, D. Oloumi, and K. Rambabu, "Design and development of an inexpensive sub-nanosecond Gaussian pulse transmitter," *IEEE transactions on microwave theory and techniques*, vol. 67, no. 9, pp. 3773–3782, 2019.