

# 文化计算方法与应用综述<sup>①</sup>

赵海英<sup>1</sup>, 贾耕云<sup>2</sup>, 潘志庚<sup>3</sup>

<sup>1</sup>(北京邮电大学世纪学院 移动媒体与文化计算北京市重点实验室, 北京 102101)

<sup>2</sup>(北京邮电大学 信息与通信工程学院, 北京 100876)

<sup>3</sup>(杭州师范大学 数字媒体与人机交互研究中心, 杭州 310025)

**摘要:** 文化计算是把计算及相关技术应用到文化领域, 探索其中的发展规律、揭示其中的内在联系或对其进行可视分析与展示。因此, 文化计算对于文化的本质挖掘和传承发扬都具有重大意义。论文对文化计算做了简单概述。首先, 从文化基因的内涵解析入手, 以文化组学为理论, 对文化计算的演化脉络和其发展历程进行总结; 介绍了一些有相似思想的文章进行文化计算理论的延展, 并对文化计算的方法进行比较。其后重点阐述了文化计算的研究内容, 包括作为信息基本模式的文化基因、核心研究对象文化量化以及可视分析的大数据应用方法, 分析各研究内容是如何对文化进行量化计算、科学分析的。最后应用案例介绍了文化计算的应用, 并对文化计算进行了总结与展望。

**关键词:** 文化计算; 文化基因; 文化组学; 文化量化; 可视化分析

## Review on the Methods and Applications in Cultural Computing

ZHAO Hai-Ying<sup>1</sup>, JIA Geng-Yun<sup>2</sup>, PAN Zhi-Geng<sup>3</sup>

<sup>1</sup>( Mobile Media and Cultural Calculation Key Laboratory of Beijing, Century College, BUPT, Beijing 102101, China )

<sup>2</sup>(School of Information and Communication Engineering, BUPT, Beijing 100876, China)

<sup>3</sup>(Digital Media & HCI Research Center, Hangzhou Normal University, Hangzhou 310025, China)

**Abstract:** Cultural Computing is the application of computation and related techniques to the field of culture, to explore its rules of development, to point out the internal connections, or to make a visual analysis and display. Therefore, it plays an important role in the data mining, protection, inheritance and development of culture. This paper makes a simple introduction to Cultural Computing. Firstly, starting from the connotation of cultural gene and using culturomics as the theory, it summarizes the development history and the evolution, introduces some articles that use similar ideas to expand the theory of Cultural Computing, and simply compares some different methods of Cultural Computing. Then, presents the main research contents, including cultural gene which plays as essential information mode, cultural quantification and visualization analysis of big data application methods, which tells how they can get a quantified and scientific research on culture. Finally, the paper uses examples to introduce Cultural Computing more clearly, and has a outlook on it.

**Key words:** cultural computing; cultural gene; culturomics; cultural quantification; visualization analysis

## 1 引言

对于文化, 迄今为止还没有统一的、公认的、令人满意的定义。笼统地说, 文化是一种社会现象, 是人们长期创造形成的产物, 同时又是一种历史现象, 是社会历史的积淀物。确切地说, 文化是凝结在物质

之中又游离于物质之外的, 能够被传承的国家或民族的历史、地理、风土人情、传统习俗、生活方式、文学艺术、行为规范、思维方式、价值观念等, 是人类之间进行交流的普遍认可的一种能够传承的意识形态。

<sup>①</sup> 基金项目:国家自然科学基金(61163044);国家社科基金重点项目(12AZD118-2,12AZD120-2);北京市科委项目(Z141100001914035) 国家科技支撑计划课题(2012BAH48F03)。

收稿时间:2015-09-23;收到修改稿时间:2015-12-02 [doi:10.15888/j.cnki.csa.005206]

文化精神产品承载民族风俗习惯以及观念等的文化内涵,其表现形式丰富多样。在传统的文化研究中,学者们通过考察有重大意义与位于时空关键点的精神产品,运用自身知识储备,通过联想、感受和推理,来挖掘相关文化产品的历史、演变、分布、特征、内涵等等。但这种方式往往工作量比较大,而且有很多内涵和规律隐藏在大量数据的深处,是不容易被发现的。

2011年,哈佛大学 Michel 研究团队在《SCIENCE》杂志上阐述了一种新思想<sup>[1]</sup>,利用海量的文本数据,做有关词汇的频率统计可以验证甚至补充某些历史事件。这种思想启示工作者可以像统计词频那样量化地研究文化现象与文化精神产品,对某种文化的演变规律、特征内涵等做出更清晰、全面的认知,并可能对其进行系统的呈现与优化。

本文研究文化计算的目的是:(1)与文化领域专家合作,在已有的文化媒体数据库的基础上,共同研究对各种文化进行数字存储的理论方法,包括影音图像,三维虚拟现实和计算机动画等;(2)研究对文化内容提取文化基因并对其进行数字描述和数字特征表达的关键技术,使其能够通过计算机重构文化形式,例如重构文化遗产,并得到有关专家的认可;(3)通过对数字文化基因的数学建模,找到其构成相应文化内容的关键技术以及各基因之间的相互关联,从而利用拓扑理论将其传承,开拓和挖掘的各种可能进行可视化。建立一个可视互动平台,帮助艺术家和专家进行有关文化内容的数字挖掘和开拓,使不同文化内容与文化载体在新时代的大数据应用背景下,发挥与时俱进的作用。中华文化源远流长博大精深。对文化进行量化研究的文化计算不仅能帮助人们更深入地了解她,而且对于中华传统文化的保护、继承与发展有着重要的意义。本文主要对文化计算的发展与现状,以及文化计算的主要内容和方法进行了综述,并介绍了与文化计算相关的一些应用案例。

## 2 发展与现状

### 2.1 文化计算相关定义

文化计算是把计算及相关技术(数据挖掘、人工智能、虚拟现实、统计学习和可视分析)应用到文化领域,探索其发展规律、提示其内在联系并对其进行量化分析与展示。

简单来说,文化计算是一套科学地采集、存储、处理、认知、呈现、优化文化的方法体系,其面向的是各种形式的文化载体如图形、图像、文字、音乐、舞蹈等。

具体说,就是采用机器学习、特征提取等手段,提取文化特征信息并预处理;然后基于这些信息对文化进行科学的量化与建模,再可视化计算、分析与展示的完整过程。在这个过程中,对提取的文化特征信息进行存储、量化、分析计算是文化计算的核心内容。

在特征信息提取处理的过程中,最重要的工作就是找出某种特定文化的文化基因,因为文化基因通常是指文化中富有生命力的、具有内在联系的遗传密码和核心元素<sup>[2]</sup>,也是体现文化差异性的重要特征。很多学者都指出了发掘文化基因并构建基因库的重要意义<sup>[3-4]</sup>,并对文化基因的实际应用作了简单探索。如在苟秉宸等的《半坡彩陶文化基因提取与设计应用研究》<sup>[5]</sup>中就分析提取了半坡彩陶文化的几种基因,并成功应用在了彩陶图案的设计中。

进一步推广,文化计算可以应用于不同文化载体,但在计算文化或文化遗产之间的基因相似度,可以把时间、空间、基因相似度作为坐标轴,建立文化或文化遗产的时空演化三维空间,并以基因相似度为主要标准,在上述三维空间中搜索文化遗产时空演化的路径,这就是文化计算的核心内涵。

### 2.2 文化计算发展与现状

在国内,虽然文化基因的概念早已有之<sup>[6]</sup>,近些年基于文化基因的文化遗产挖掘、保护也越来越受到关注<sup>[7]</sup>,但是这些研究仍然是偏人文的、偏论述性,没有深入地与文化时空演化、信息挖掘技术与可视化分析结合。

日本京都大学的 Tosa Naoko 在 2005 年前后提出了文化计算的概念<sup>[8]</sup>,在此之前,日本国立情报研究所的 Noriko Kando 教授和 Jun Adachi 教授就对文化遗产库的建立做了有益的分析与尝试,并帮助建立了“文化遗产在线”网站<sup>[9]</sup>,大大推动了日本文化数字化的进程。

Tosa Naoko 教授认为文化的数字化不应只停留在存储、保护上,她提出文化计算是通过数字化技术捕捉并多方位地呈现某种文化,进一步分析用户在特定文化情景和体验下的感受、情感乃至潜意识的思考,从而能够呈现或感受到文化的精髓。相关研究内容集

中在交互感知、虚拟现实、人的情感意识分析等方面,关注的是文化所传达的感受和意境,对文化本身的科学物质特征关注较少。

2011年哈佛大学Michel研究组提出了一种新的视角和方法,通过对图书馆中书籍资料的大数据量化分析推动了各领域的量化研究.通过统计引用 Michel研究组论文《Quantitative Analysis of Culture Using Millions of Digitized Books》的情况,可以部分地反映相应思想方法的研究与应用情况(2015年的引用截止到7月底),如图1所示。



图1 对《Quantitative Analysis of Culture Using Millions of Digitized Books》的引用数

Michel等人将他们的工作概括为“文化组学”。郭崇慧等在《文化组学研究综述》<sup>[10]</sup>中提出,文化组学就是“通过获取历史文字资料中单词和短语随着时间推移的使用频率,形成可视化的使用频率波动图,依此量化人类文化演变轨迹”。

文化组学是在文化计算中量化分析文化的一种重要手段,由于语言文字内容相较于其它文化形式具有语义含义清晰、存储方便、有较好的统计规律等特点,因此对它的研究起步早而且全面,其思想方法在许多学科领域都有被涉及.如Armstrong.D通过对医学文献数据的分析得出新的慢性病蔓延的原因<sup>[11]</sup>,Soper.DS等人通过对系统科学顶级期刊的有关内容的频率统计,分析哪些是该领域的核心定理公式<sup>[12]</sup>,此外,还有很多学者逐渐发展了统计、量化等思想,实际上也逐渐超出了文化组学概念的范畴,开始探讨更深入的应用.表1简单列出了几个相关的研究工作。

表1 运用文化组学思想的相关研究

时间	作者	研究内容
2012	Weingart,S; Jorgensen,J	计算分析欧洲童话故事中的 人物 <sup>[13]</sup>
2013	Amancio,DR ; Altmann, EG; Rybski, D ; Oliveira, ON ; Costa_LD	研究如何鉴别一份未知的文本 属于哪种语言,并应用在伏尼 契手稿上 <sup>[14]</sup>
2013	Tangherlini_ TR; Leonard_P	提出了一种子语料库主题建模 方式,可以发掘未理解文本中 的有意义段落 <sup>[15]</sup>
2013	Oishi,S; Graham,J; Kesebir,S; Galinha,IC	综合运用各类词典与谷歌 ngram,挖掘不同历史时期与文 化中“幸福”概念的关系 <sup>[16]</sup>
2014	Liu,Y ; Li,XF ; Te,RG ; Pan,C; Zang,XB	提出一种音乐基因提取的方法, 并用于按历史时期分类 <sup>[17]</sup>
2014	Popa, T ; Rebedea, T ; Chiru, C	令特定的词语、主题与特定的 历史时期相匹配的方法研究 <sup>[18]</sup>
2014	Acerbi, A ; Bentley, RA	文化传播中的偏见对流行文化 元素形成的作用 <sup>[19]</sup>

在文化组学基础上提出的文化计算更加注重对多种不同形式文化本身的科学分析而不再局限于文本资料,提出了包括文化基因和元数据等在内的一系列分析方法.文化的存在形式是多样且复杂的,这一方面反映了文化遗产的丰富和人们文化生活的多彩,另一方面却给文化计算带来了挑战,目前对文化计算的相关研究集中在几个方面:(1)文化基因的提取与简单重构应用;(2)文化基因和文化元数据的科学存储模型研究;(3)文化内涵与文化发展规律本质的简单挖掘。

### 2.3 社会计算

社会计算的概念近些年也受到很多关注<sup>[20][25]</sup>,其主要根植于海量的社会与网络媒体数据以及发达的计算机数据分析能力,是数据密集型的的研究<sup>[21]</sup>.它的主要应用在于对各类社会媒体的数据分析(如网络数据质量分析<sup>[22]</sup>)与数据挖掘(如Twitter事件检测<sup>[23]</sup>、情感分析<sup>[24]</sup>等),从中发现模式、规律与知识。

显然社会计算与文化计算有很大的相似之处,而其最主要也是最核心的不同在于所分析对象的不同,前者所分析的往往是网络社交媒体数据,后者则往往是文化载体.因此,文化计算既要从社会计算中借鉴相应的方法,又需要探索自己特有的研究规律与模式。

## 3 相关工作—文化计算方法比较

### 3.1 横向比较

这里简单比较了两类文化计算的方法,表2比较了文化基因库建设的两种方式,表3比较了文化量化的两种方式。

表2 文化基因库建设的方法比较

文化基因库构建的方法	优点	缺点
计算机智能提取与构建	能够处理大量数据,快捷高效节省人力,发展潜力大	准确度不够高,有些特征计算机无法直接提取
人工提取与构建	特征提取效果好,分类明确清晰,不易出现错误	耗费人力,用时长,处理数据量小

在文化基因方面,很多国内学者做了人工提取构建一些小规模文化基因库的工作<sup>[26-28]</sup>来尝试分析和保护特定的文化遗产,或用于存储<sup>[29]</sup>。而在文化计算中,研究者尽可能的把数字化手段引入到这个过程中来,以提高研究的效率和质量。

表3 文化量化分析的方法比较

文化量化处理的方法	优点	缺点
文化组学	在文本数据分析中应用广泛,容易与文献计量学结合,数据库构建相对容易	只能覆盖小部分文化形式,分析方法单一,难以挖掘本质规律与模式
基于元数据(包括文化基因)的文化量化	对各种文化形式都可适用,可以深入发掘文化本质,可以有效促进文化保护与发展,发展潜力大	前期处理困难、基因库构建相对较难,分析方法较为复杂,

文化组学在一开始受到了广泛关注,但加州大学伯克利分校的语言学家 Geoffrey Nunberg 认为“将这种方法称为“文化组学”是相当自大傲慢的,因为新方法中的绝大多数分析几乎都是相当粗糙的”<sup>[30]</sup>,基于元数据的文化量化在一定程度上弥补了它的缺陷,同时也具有较大的发展空间,但对二者来说,新的数据源和分析方法都有待进一步挖掘。

### 3.2 纵向比较

在文化组学基础上发展起来的文化计算与 Tosa Naoko 等提出的文化计算相比有很大不同,它们虽然都是把数字、信息技术与文化相结合,但是前者注重对文化本身的量化分析,通过科学的存储与建模分析,可以达到发掘文化本质,找出文化规律的目的。而后者,则更多的关注了“人”的方面,通过文化的多维展

现以及与人的互动,可以从人的层面与角度理解文化精髓,能表达出文化中的内涵与意境。二者都是信息化数字化时代,对文化与信息技术深度结合的有益尝试,对于保护、发展文化都有重要意义。

## 4 文化计算核心研究内容

### 4.1 基于文化基因的文化计算

文化基因的提取与基因库的构建是文化计算核心步骤中的第一步,是文化可以被量化、被分析的重要依据。

文化基因的提取实际上是针对文化背后蕴含的特征和文化模式的挖掘和发现,是对进行更多的定量综合分析做准备。刘沛林在研究聚落景观基因时提出了一些确认原则<sup>[31]</sup>,主要是要求提取出的文化基因应当体现文化的独特性。虽然当前文化基因的提取已有不少进展,但对于如习俗礼仪、音乐舞蹈、图案壁画等非结构化的文化遗产,如何发掘其背后蕴藏的特征与文化模式仍然是非常具有挑战性的难题。

另一方面,基因库的构建是科学、便捷地使用文化基因的前提。一般来说,文化基因包含多个属性及关系,同时还可能有多种层次,这使得不同形式不同种类文化的文化基因可能有不同的存储方式与存储结构,如何根据基因特征和使用方式科学建设文化基因库也是目前面临的重要课题。基于文化基因演化的时空一致性约束确定基因的组合变异规则,在此基础上可根据已有基因生成新的模式,这是基于文化基因的文化计算核心目标。

### 4.2 基于文化量化的文化计算

文化量化,即是对文化的量化处理,这是文化计算中最关键的一步,同时也与可视化分析紧密相连,主要研究内容包括:

#### (1) 文化组学

依前文所述,文化组学即是文化量化分析的一种重要手段,它是通过词汇频率与文化发展的内在联系,通过统计其规律来量化反映文化的演变。在 Popa, T 等人的工作中,不仅使用了词汇,还使用了文本主题信息,这也充分说明了文化组学的语料库是丰富多样的,从而通过合适的基因库构建,文化组学的方法也可以应用在图像、音乐等文化表现形式中。

#### (2) 元数据量化

目前国际上已有不少成熟的元数据标准,如

DC<sup>[32]</sup>、CDWA<sup>[33]</sup>等。在文化基因的提取之后,可以使用文化基因与一系列元数据(可以认为文化基因是一个特殊的元数据)在元数据多维空间上进行量化,这样就可以把表达文化的信息在一定程度上结构化。目前这些文化特征的获取仍然有一定的难度,一般依靠底层特征到高层(文化、语义)特征的映射,或是利用大数据挖掘技术,通过特征深度学习,进一步训练确定文化特征。

需要注意的是,文化基因与元数据这些文化的特征属性不可能表达出全部信息,故而要实现分析的目的,元数据的选择或设计是至关重要的,也就是选择在内容上为主要的,在结构上居于主体的,在演变过程具有一定稳定性和继承性的成分。因此能够用它们大致地呈现文化的样貌并且有利于分析计算。Noriko Kando 教授和 Jun Adachi 教授从文化遗产存储的角度指出了元数据的重要作用,包括可以使检索更有效、可以反映相似或属于同一范畴的文化遗产的联系等<sup>[9]</sup>。

从另一个角度考虑,包含文化基因在内的这些元数据构成了表达文化的一组基,在数学上意义上,它应该被要求是完备的和正交的。而要保证这两点同时满足是很困难,因此研究文化时常常有所侧重,可能只能一定程度上,或在一定的方面满足即可。

### (3)特征关系的量化计算(相似性度量)

对于文化特征(文化基因与元数据),一个特征的不同属性值之间的关系是最重要的性质之一。特征关系的量化计算,即是对特征之间相似程度的量化估计。一方面,不同种类文化的基因甚至元数据内涵差距较大,因此需要不同的估计方法;另一方面,不同的文化特征,量化难易也不同,例如,元数据中的“时间”项就是可直接量化计算的简单项,而文化基因则一般比较复杂,例如形态颜色等,其相似度的量化常常需以情况而定,还可能参考其他数据。

文化特征的相似度量在分析文化的发展进程以及文化随地域分布的变迁中有重要作用,通过分析不同时间或不同地域在文化特征上相似性的大小,可以模拟文化发展、迁移的过程,并能够发现其中的规律。

## 4.3 基于可视分析的文化计算

人类从外界获得的信息约有 80%以上来自于视觉系统,当大数据以直观的可视化的图形形式展示在分析者面前时,分析者往往能够一眼洞悉数据背后隐藏的信息并转化知识以及智慧。大数据可视分析旨在

利用计算机自动化分析能力的同时,充分挖掘人对于可视化信息的认知能力优势,将人、机的各自强项进行有机融合,借助人机交互式分析方法和交互技术,辅助人们更为直观和高效地洞悉大数据背后的信息、知识与智慧<sup>[34]</sup>。

目前,文化计算中的可视化分析方法主要用于帮助研究人员更方便地分析海量文化的关系与变化,而利用图形学的三维场景建模等技术还可以直观的展示相关的文化产品,从而引起人们的兴趣,促进文化的广泛传播。可视化分析方法的选择依赖于分析数据的形式和分析的目的。例如当分析文化演化、分布规律时,可以根据文化基因与元数据在合适的时空坐标上对文化进行表达,进而直观地进行分析。当把线性文化遗产作为研究对象时,利用影像技术、数字化等科技手段对文化遗迹的色彩、几何特征进行提取与度量,并在此基础上,通过一个统一的时间、空间和文化特征三维空间中研究相适应的可视分析技术,以便于搭建一个高效的沉浸式可视分析系统,提高文化遗产的演化规律和传播路径研究的效率。

文化计算的一个重要方面是揭示文化变迁、发展等的演化特点和规律,由此为现今文化的发展提供历史借鉴。因此,需要在繁杂的文化数据中根据文化特征探索其演化的时空关联性。但是面对时空跨度大、时空耦合强、形态多样、数量庞大等特点的文化内容,如何在时空一致性要求下进行文化特征的时空关联性是一个非常困难的问题。

## 5 典型案例—文化计算应用案例篇

### 5.1 中国传统聚落景观基因图谱的构建与应用研究<sup>[30]</sup>

#### (1)传统聚落景观基因的表达和提取

研究了聚落景观基因的主要表现形式或者说内容,包括布局、房屋主要部分造型、形态、装饰和用材等。

提出确认传统聚落景观基因的原则,即内在唯一性原则、外在唯一性原则、局部唯一性原则和总体优势性原则。

#### (2)传统聚落景观基因图谱的建立

分为平面图谱和立体空间图谱两类,对每一类又分别构建了跨区域和同区域的基因图谱。提出借助生物学的基因图谱序列逻辑和人类学的谱系排列逻辑建构。

使用了多种示意图表达基因图谱。在传统城镇景

观基因的表达上,创造了“胞—链—形”的图示表达方式,深入直观地展示了城镇景观的本质和内在联系

### (3)传统聚落景观的区系和特征

介绍分析了与聚落景观区系相关的建筑区系研究,提出了一些中国传统聚落景观区划的原则和存在的某些技术问题.并按照“大区—区—亚区”的三级划分法,形成了中国传统聚落景观区划的初步方案.

认为聚落景观基因是判断传统聚落景观区系的核心要素,提出并分析了影响传统聚落景观基因识别的几个基本要素,包括理要素、生态要素、美学要素、环境要素、文化要素、时序要素.

### (4)传统聚落景观的保护和利用

充分利用了地理信息系统 GIS,分析了 GIS 在景观基因识别与提取、聚落保护与管理中的重要作用.

鉴于当前文化旅游地规划的不规范性以及旅游规划理论匮乏的现实,提出了“景观信息链”理论,将代表地方特色的传统聚落的历史文化信息(即景观基因)挖掘整理出来,按照一定的范式进行筛选和提炼,然后通过景观再现和景点组合等方式有效表达出来,使文化旅游地的地方感强、特色鲜明、个性明显.

## 5.2 用于按时期分类的音乐基因提取<sup>[17]</sup>

指出当前研究人员提出的音乐特征忽略旋律音乐的整体性.提出音乐基因的概念.音乐基因应当是有意义的最小的语法单元,它把整体性考虑在内.音乐基因可以被理解为一个函数,他把音符的组合映射到一些更为抽象的概念上去.

对比了两类音乐存储格式,并使用 Sheet-MusicXML 作为音乐基因提取的数据源,认为其优势在于(1)对于半结构化的音乐数据有更清晰的结构;(2)对音乐基因的提取来说它更具有数据的完整性.

音乐基因设定为以音程信息为主,即音符之间的音高关系,音乐基因提取的步骤为:

(1)对给定的音乐,按一定长度截取音高信息,并进一步计算音程.

(2)将音程信息集按一定规则分组,规则包含指定的音程集的差异阈值.

(3)计算每一个组的频率,并与给定的最小支持度作比较.

最后运用 SVM 分类,研究三个参数(最小支持度、差异阈值、基因长度)对分类结果的影响,并与其它音乐特征的效果作比较.分析得出考虑音乐整体性的音

乐基因特征在音乐按时期分类时会更加有效.

## 5.3 半坡彩陶文化基因提取与设计应用研究<sup>[5]</sup>

由生物学中基因的作用,提出半坡彩陶文化产品的设计过程就是以半坡彩陶文化基因为基础的半坡彩陶文化元素生长发育过程.

半坡彩陶文化基因的提取:用层次结构图来确认影响半坡彩陶文化风格的基因,指出图案基因、色彩基因、形态基因是文化产品设计中最典型的基因.

基因的提取基本采用人工识别干预的方式,并提出了文化基因的显性基因与隐性基因的概念,指出采集图案、提炼某种图案基因的一般核心特征的过程处理的是显性图案基因,而对显性基因进行继承和改进,从而演绎出的更多同类图案则属于隐性基因.作者认为这种理念方法有助于丰富基因库.

指出应当对文化基因的优劣进行分析,重点保留有利于传承的优秀基因.构建了简单的半坡彩陶基因库,基因库包含基因的属性、描述和存储信息.应用设计软件系统,以半坡鱼调料瓶为例,对使用基因库进行半坡文化产品设计进行了简单介绍,图 2 为其基因库系统的截图.



图 2 半坡彩陶文化图案基因库

## 5.4 中华文化基因抽取与特征建模探索<sup>[2]</sup>

文章以北师大的“中国文化符号调查报告”引出对抽取中华文化基因的难度的思考与分析,并提出发展与繁荣中国文化应遵循的步骤,即:(1)找到可以永远传承和坚守的中华文化遗传基因和核心元素;(2)应该适应受众不断变化的心理需求和审美标准的传播方式;(3)让传统文化“反客为主”、焕发活力并在新技术与当下文化表达的符号中发挥重要作用.

提出了文化基因抽取与特征建模的基本路径.首先应当利用数字化技术加强对文化遗产的贮存保护,其次应当充分利用文化组学,量化分析中华文明的

DNA, 构建文化基因库与文明基因图谱, 其中基本的学术研究路径应当包括对词汇、图像、音响、语义的分析. 最后, 应对量化分析的结果与以实证考察为基础的民间文化、考古文化与历史文献相互结合、相互印证, 不但可以查缺补漏, 还能实现中华文化的深度挖掘.

作者最后提出了建设“文化遗产 DNA 在线”传播工程的目标, 认为该工程将有利于传统文化教育与研究, 提高中华文明基因的利用率与接近性; 使文化遗产从地域的限制中独立出来, 更易保护; 能促进中华文明与其他文明之间的结合、对比与修复, 进而加强不同文化体之间的相互理解.

### 5.5 文化计算应用案例分析

表4 文化计算应用案例分析

案例	方法使用	特点分析
中国传统聚落景观基因图谱的构建与应用研究	文化基因、可视化分析	景观基因的概念与多学科(历史地理学、聚落类型学等)方法融合, 与可视化方法结合, 得到了聚落形态、分布与演变等的科学结果, 并能够运用于旅游等领域. 是对景观基因的多角度综合利用
用于按时期分类的音乐基因提取	文化基因、自动化的量化分析	文化基因概念在音乐自动化研究领域的成功应用. 利用音乐基因, 基于 XML 存储与有监督的 SVM 分类实现了一整套的自动化提取与研究系统, 考虑全面、智能化程度高.
半坡彩陶文化基因提取与设计应用研究	文化基因	文化基因在工业设计领域的成功应用, 通过对半坡彩陶文化基因的提取-存储-组合应用, 借助计算机实现了交互式的半坡彩陶设计
中华文化基因抽取与特征建模探索	文化基因、文化组学	对中华文化基因相关工作的系统性探讨, 提出了一些在文化基因基础上发展中华文化的有益的、前瞻性的建议, 对整体性的文化工作有很强的指导性.

## 6 总结与展望

文化计算把数学的、科学的思想应用到历史学等人文学科中, 一方面反映了人文社科发展的数字化进程, 另一方面也反映了当代学科融合已经向深处发展. 科学和人文之间不再是难以逾越的鸿沟, 越来越多的人也意识到各学科的思想、方法和应用是没有界限的.

过去人们之所以没有用数学的、量化的眼光去看待、分析文化, 除了观念的影响外, 更多的还是科学水平的限制, 学者们难以搜集、保存大量的文化样本及文化产品信息, 同时也缺乏从文化表达中提取特征信息并进行分析和展示的有效手段. 但是随着人们思想观念的转变和技术水平的进步, 这些问题都已得到或将得到解决. 信息时代、大数据时代的来临, 使得获取、存储、调动海量的数据成为可能. 而计算机视觉和图形技术的发展、成熟以及数据挖掘技术的快速发展, 使得人们面对多样的文化及其产品时不再只能依赖专业学者的知识和经验, 而是可以依靠先进的技术去更广泛、更深入地研究文化、传播文化.

文化计算是文化学者的福音, 更是传统文化爱好者、继承者的福音, 文化计算能够帮助保护乃至创新传统文化, 促进多元文化的繁荣与发展.

当然, 文化计算目前还有很多不足, 发展也不够完善. 全面系统的基因库的构建仍然缺乏, 系统的多样化的文化量化与计算的方法体系仍然没有完全形成. 同时信息时代既是机遇也是挑战, 它既给人们提供了计算文化的数据与技术, 又使得文化的发展面临激烈变革. 在信息样式繁杂内容良莠不齐、信息变化剧烈时效性强的今天, 文化计算虽然面临不少挑战, 但一定会成为研究、发展人类文化的强力助手, 使得文化在数字化时代取得更大的发展, 展现出更强的魅力.

结束语 本文提出了一种运用计算机等技术科学分析文化及其产品的思想, 介绍了当前文化计算的研究内容、研究方法与研究案例, 认为文化计算能够对探索文化的内在模式和规律提供帮助. 并对其进行了总结与展望.

### 参考文献

- 1 Jean-Baptiste M, Shen YK, Presser AA, Adrian V, Matthew K G, Pickett JP, Dale H, Clancy D, Norvig P, Orwant J, Pinker S, Nowak MA, Aiden EL. Quantitative analysis of culture using millions of digitized books. *Science*, 2010, 331:6014.
- 2 邵培仁, 林群. 中华文化基因抽取与特征建模探索. *徐州师范大学学报(哲学社会科学版)*, 2012, 02:107-111.
- 3 李心峰. 人类艺术基因库与中国少数民族艺术. *民族艺术*, 2004, 2:47-50.
- 4 吴良林, 周永章, 陈子燊, 丁健. 桂西山区原生态旅游资源及其可持续利用研究. *生态经济*, 2007, 2:102-105.
- 5 苟秉宸, 于辉, 李振方, 陈丽伶, 王伟伟. 半坡彩陶文化基因提

- 取与设计应用研究.西北工业大学学报(社会科学版),2011,4:66-69,104.
- 6 蔡镇楚.中国诗话与民族文化性格.中国文学研究,1987,3:11-16.
- 7 王东.中华文明的文化基因与现代传承(专题讨论) 中华文明的五次辉煌与文化基因中的五大核心理念.河北学刊,2003,5:130-134,147.
- 8 Tosa N, Matsuoka S, Ellis B, Ueda H, Nakatsu R. Cultural computing with context-aware application. In: Kishino F, et al. Eds. ICEC 2005, LNCS 3711, 2005. 13-23.
- 9 Kando N, Adachi J. Cultural heritage online: Information access across heterogeneous cultural heritage. Japan.kc.tsukuba.ac.jp
- 10 郭崇慧,魏伟,任晓玲.文化组学研究综述.情报学报,2014,33(7):765-774.
- 11 David A. Chronic illness: A revisionist account. *Sociol Health Illn*, 2014, 361.
- 12 Soper DS, Turel O, Nitza G. The intellectual core of the IS field: A systematic exploration of theories in our top journals. 2014 47th Annual Hawaii International Conference on System Sciences. Univ Hawaii, Shidler Coll Business. IEEE Comp Soc. 2014. 6.
- 13 Weingart S, Jorgensen J. Computational analysis of the body in european fair7 tales. *Literary and Linguistic Computing*. 2012: 1-13.
- 14 Amancio DR, Altmann EG, Rybski D, et al. Probing the statistical properties of unknown texts: application to the Voynich Manuscript. *PloS One*, 2013, 87.
- 15 Tangherlini TR, Leonard P. Trawling in the sea of the great unread: Sub-corpus topic modeling and humanities research. *Poetics*, 2013, 6: 725-749.
- 16 Shigehiro O, Jesse G, Selin K, Iolanda Costa G. Concepts of happiness across time and cultures. *Personality & Social Psychology Bulletin*, 2013, 395.
- 17 Liu Y, Li XF, Te RG, et al. Extracting music genes for era classification. *Expert Systems With Applications*, 2014.
- 18 Popa T, Rebedea T, Chiru C. Detecting and describing historical periods in a large corpora. *IEEE International Conference on Tools with Artificial Intelligence (ICTAI)*. 2014. 11.
- 19 Acerbi A, Bentley RA. Biases in cultural transmission shape the turnover of popular traits. *Evolution and Human Behavior*, 2014.
- 20 孟小峰,李勇,祝建华.社会计算:大数据时代的机遇与挑战. *计算机研究与发展*,2013,12:2483-2491.
- 21 Tang L, Liu H. 文益民,闭应洲,译.社会计算:社区发现和社会媒体挖掘.北京:机械工业出版社,2013.
- 22 韩晓晖.Web 社交媒体中信息的质量评价及应用研究[学位论文].济南:山东大学,2012.
- 23 Weng J, Lee BS. Event detection in Twitter. *Proc. of the 5th Int AAAI Conf on Weblogs and Social Media (ICWSM'11)*. Menlo Park, CA. AAAI. 2011. 401-408.
- 24 Barbosa L, Feng J. Robust sentiment detection on Twitter from biased and noisy data. *Proc. of the 23rd Int Conf on Computational linguistics (COLING'10)*. Stroudsburg, PA. ACL. 2010. 36-44.
- 25 Lazer D, et al. Computational social science. *Science*, 2009, 323, (5915): 721-723.
- 26 曹帅强,邓运员.非物质文化遗产景观基因的挖掘及其意象特征——以湖南省为例. *经济地理*,2014,11:185-192.
- 27 刘沛林.古村落文化景观的基因表达与景观识别. *衡阳师范学院学报(社会科学)*,2003,4:1-8.
- 28 赵鹤龄,王军,袁中金,马涛.文化基因的谱系图构建与传承路径研究——以古滇国文化基因为例. *现代城市研究*, 2014,5:90-97.
- 29 特日根,李雄飞.基于音乐基因的乐谱存储模型. *中国科学:信息科学*,2014,7:946-950.
- 30 王丹红.文化组学:用数学方法分析文化文化演变. *科学时报*. 2011-1-18.
- 31 刘沛林.中国传统聚落景观基因图谱的构建与应用研究. 北京大学,2011.
- 32 DCMI Usage Board. properties in the elements 1.1 namespace. <http://dublincore.org/documents/dcmi-terms/>. [2015-11-1].
- 33 Categories for the Description of works of Art(CDWA). [http://www.getty.edu/research/publications/electronic\\_publications/cdwa/definitions.html](http://www.getty.edu/research/publications/electronic_publications/cdwa/definitions.html). [2015-11-1].
- 34 任磊,杜一,马帅,张小龙,戴国忠.大数据可视分析综述. *软件学报*,2014,9:1909-1936.