感知性与和谐度引导的多图表协同配色优化

王晖, 胡瑞珍*

(深圳大学可视计算研究中心 深圳 518061)
(ruizhen.hu@gmail.com)

摘要: 针对在可视化过程中出现的具有相同类别对应关系的多个图表需要保持配色一致的问题, 提出多图表协同配色优化, 从感知性与和谐度 2 个方面提出感知性与和谐度引导的多图表协同配色优化方法. 给定调色板和多图表数据后, 对调色板的质量进行评估, 若给定的调色板质量较高, 则对其颜色分配方案直接进行优化; 否则, 先对调色板进行优化, 再对其颜色分配方案进行优化. 多项用户调研结果证明, 该方法能够在保证多图表配色统一的前提下提升配色结果的质量.

关键词: 配色优化; 感知性; 和谐度; 遗传算法

中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2021.18639

Perception and Harmony Guided Color Assignment Optimization for Multi-Charts

Wang Hui and Hu Ruizhen*

(Visual Computing Research Center, Shenzhen University, Shenzhen 518061)

Abstract: In order to ensure the visual effect and unity of color mapping of multi-charts among the data with the same category, it is proposed to study the color assignment optimization problems on multi-charts, which introduces a score defined based on perception and harmony to guide the joint color assignment optimization. When given the data plotted with multi-charts and the color palette, first the quality of the color palette is evaluated. If the color palette is of high quality, the proposed method will directly optimize the color assignment. Otherwise, the color palette needs to be optimized first. Several user studies are conducted to show that the proposed method can ensure the unity of multi-charts color mapping and improve the quality of visualization results.

Key words: color assignment optimization; perception; harmony; genetic algorithm

在可视化过程中, 为了增强用户对数据的认知和理解, 设计人员会通过调整不同的视觉通道改善最后的可视化结果, 如大小、位置还有颜色等. 而在这些视觉通道中, 颜色作为影响视觉效果的主要因素, 其优化效果非常值得研究, 且已经有大量这方面的工作.

对于颜色的优化效果, 可以从感知性与和谐度 2 个方面进行刻画, 即优化的可视化图表需要增

强不同数据间的辨识程度, 同时保证图表整体配色的和谐相融性. 当给定调色板与数据后, 若按照调色板中颜色与数据类别按顺序对应的关系进行配色, 就会导致图表出现一些较小的数据被分配到不易感知的颜色, 而一些较大的数据被分配到视觉效果不佳的颜色, 使最终的可视化结果感知性与和谐度较差. 因此, 需要同时对调色板本身以及调色板的颜色分配进行优化.

收稿日期: 2020-09-05; 修回日期: 2021-01-05. **基金项目:** 国家自然科学基金(61872250). **王晖**(1992—), 男, 硕士研究生, 主要研究方向为可视化; **胡瑞珍**(1988—), 女, 博士, 助理教授, 硕士生导师, CCF 高级会员, 论文通讯作者, 主要研究方向为计算机图形学、可视化.

相对于先前工作只研究单张图表,本文主要发现在实际应用过程中,由于数据的维度大,需要用多张图表展现不同维度的信息,而这些图表之间存在很强的数据对应关系,因此需要对多图表协同配色进行优化,以达到统一的颜色分配方案和良好的视觉效果。另外,与多图表类似的还有同一图表多布局的情况。例如,对于一些图表,存在按数据输入顺序以及数据大小顺序进行布局的 2 种形式,这种情况同样需要协同优化。

因此,提出了感知性与和谐度引导的多图表协同配色优化方法,该方法具有很好的普适性,并且对调色板的颜色数量没有限制。在给定调色板和上述数据之后,该方法会先对调色板的质量进行评估,若给定的调色板质量较高,则会对其颜色分配方案直接进行优化,旨在提升可视化结果的感知性与和谐度。若给定的调色板质量较低,则会先对调色板进行优化,提高调色板整体的感知性与和谐度,然后再对其颜色分配方案进行优化。而在颜色分配优化的过程中,主要是通过提取各图表相邻色块,建立起多图表色块的几何约束关系进行引导。为了验证所提出方法的有效性,本文开展了多项用户调研,结果显示本文所提出的调色板优化方法和调色板的颜色分配优化方法均能有效提高可视化结果的感知性与和谐度,且由于问题复杂度高,此种异步优化效果要优于同步优化,即同时优化调色板和颜色分配。

1 相关工作

目前关于多图表以及同一图表多布局协同配色的研究尚少,所以本文从调色板和颜色 2 个方面对相关工作进行论述。

1.1 调色板

对于具有类别的数据可视化,Trumbo^[1]和 Wang 等^[2]认为,以最大限度的区分度创建一个分类调色板是一项要求很高的任务。因此,大多数可视化用户会尽可能地避免从头创建调色板,而是会直接使用 ColorBrewer^[3]或 Colorgorical^[4]等提供的调色板。本文也是如此,选用的调色板来源于 ColorBrewer^[3]以及 Tableau^①等常见的可视化工具。

在选定调色板后还可以从多方面对其进行更进一步的优化,如颜色和谐^[2]、能量消耗^[5]、类别可见度^[6]以及感知距离^[7]等。Lee 等^[6]提出的类别可

见度是通过每个类别的感知密度来定义的,而感知密度则是通过计算所有点的凸显度总和得到的。基于上述思想,他们提出了一种通过优化给定调色板的感知强度,从而更好地揭示数据分布结构的方法,主要用于地图结构的可视化。Fang 等^[7]则是通过最大化调色板颜色之间的感知距离优化调色板,并允许用户加入一定的限制。

在给定调色板和类别数据后,调色板颜色分配的目的是为了给每个类别的数据配上唯一的颜色标识。Wang 等^[8]对散点图的配色进行优化,在该方法中,颜色的选取属于局部选择。Kim 等^[9]则对 2D 图片模板配色问题进行研究,在该方法中,对调色板的颜色数量有限制要求。Lin 等^[10]提出了一种选择语义相关颜色的方法并证明方法的有效性。Setlur 等^[11]为了进一步提高 Lin 等^[10]方法的效果,采用谷歌 n -gram 的颜色名称频率共现的测量方法。

本文改进了前人的方法,使其能够适用于多图表的协同优化,并且对调色板的颜色数量没有限制。

1.2 颜色

本文涉及颜色的相关研究主要分为 2 大类:颜色之间的和谐度,即颜色相互搭配是否足够好看;颜色之间的感知性差异,即颜色相互之间是否容易区分。

Paterson^[12]认为,通常情况下,当颜色并置在一起产生一种令人满意的统一或平衡时,这些颜色就被认为是和谐的。换句话说,当一些颜色放在一起能够产生令人愉快的效应时,可被认为是和谐的^[13]。

有很多研究人员或艺术家尝试去建立一个基本的颜色和谐规则,最早的颜色和谐规则如下:当颜色色调相同,但饱和度或明度不同时;当颜色相似或互为补色时,即被认为是和谐的。此外,许多研究人员还尝试通过分析实验或调研结果,去构建一个良好的颜色和谐模型。Papachristos 等^[14]定义了一个模型,预测给定配色方案的情感与审美价值。Ou 等^[15]开发了双色组合和谐度的计算模型。Szabo 等^[16]在 CIECAM02 颜色模型的基础上开发了类似的模型,该模型除了可以用于双色组合外,还可以用于三色组合的评价。

除了颜色的和谐度之外,大多数情况下,研究人员还希望能够衡量颜色之间的感知性差异。为了计算颜色的感知性差异,会将颜色统一到相同的颜色空间下,这样颜色之间的感知性差异大小

① <https://www.tableau.com>

就会与颜色空间下的几何距离相对应。

在标准空间下, 常见的衡量方法有 CIE76, CIE94 和 CIEDE2000^[17]等, 其中 CIEDE2000 的效果要优于其他方法。除了上述方法外, 有时还可以使用亮度差异、色调差异等基本的颜色属性进行简单的衡量。

本文采用 CIEDE2000 以及 Ou 等^[18]的颜色模型等现有方法计算颜色的感知性与和谐度, 并未提出新的衡量方法。

2 本文方法概述

本文基于颜色的感知性与和谐度 2 种度量, 提出了一种对多图表协同配色优化的方法, 其流程如图 1 所示。

图1 本文方法流程

当给定调色板和多图表数据后, 先对调色板的质量进行优化, 同时从给定的数据中提取相应的几何约束, 最后将这些约束与调色板结合在一起, 对颜色分配方案进行优化, 得到最终的可视化结果。值得注意的是, 在给定的调色板质量已经很高的情况下, 可直接进行颜色分配方案的优化。

根据流程, 本文的优化目标包括调色板的优

化和调色板颜色分配方案的优化。针对提出的 2 个优化目标, 均采用遗传算进行求解。通过调整相应的染色体形式、目标方程、交叉函数以及变异函数等, 使遗传算法能够适用于不同的问题, 具体细节会在后面的章节进行说明。

3 调色板的优化

给定调色板后, 最终的可视化结果会受到调色板本身的影响。如果给定的调色板本身质量不够理想, 那么无论怎样优化颜色分配方案, 最终都不会取得很好的效果。所以针对这种情况, 可以对调色板进行优化之后再使用。通过改变调色板中颜色的基本属性, 从而提高调色板的质量。

3.1 目标方程

在已有调色板优化工作^[5-7]的基础上, 本文将感知性与和谐度引入目标方程。

定义 $C = \{C_1, C_2, \dots, C_n\}$ 为具有 n 种颜色的调色板, $S(C_p, C_q)$ 为 2 种颜色组合的分数, $S_{\min}(C)$ 为调色板中所有颜色组合中的最低分数, 即

$$S(C_p, C_q) = H(C_p, C_q) + \alpha D(C_p, C_q);$$

$$S_{\min}(C) = \min[S(C_p, C_q)], \forall p, q (p \neq q).$$

其中, $H(C_p, C_q)$ 用于计算颜色组合的和谐度; $D(C_p, C_q)$ 用于计算颜色之间的感知性差异。借鉴 Fang 等^[7]的优化思想, 本文的目标方程是 $S_{\min}(C)$, 通过最大化 $S_{\min}(C)$ 的值, 得到最终的优化结果。

使用 Ou 等^[18]提供的颜色和和谐度模型计算和谐度, 即

$$H(C_p, C_q) = H_{\Delta H} + H_{\Delta C} + H_{\Delta L} + H_{Lsum};$$

$$H_{\Delta H} = -0.7 \tanh(-0.7 + 0.04 H_{\Delta pq});$$

$$H_{\Delta C} = -0.3 \tanh(-1.1 + 0.05 C_{\Delta pq});$$

$$H_{\Delta L} = 0.4 \tanh(-1.1 + 0.05 L_{\Delta pq});$$

$$H_{Lsum} = 0.3 + 0.6 \tanh[-4.2 + 0.028(L_p + L_q)].$$

其中, $H_{\Delta pq}$, $C_{\Delta pq}$ 和 $L_{\Delta pq}$ 分别对应 CIELAB 空间下的色调差异、饱和度差异和亮度差异。

该模型具有较好的普适性。由于人们对于颜色的使用习惯受到很多因素的影响, 不同地区的人对颜色的喜好是不同的, 而之前关于颜色和谐度的工作, 都是基于大量的用户调查数据所得到的, 这些数据源各有不同, 所以不同的模型之间效果也各有差异。Ou 等^[18]收集了不同模型的数据集, 并从多个方面对这些数据集进行分析, 得到一

个新的颜色和谐度模型,该模型也因此具有较好的普适性.

使用 CIEDE2000^[17] 距离计算感知性差异 $D(C_p, C_q)$.

Ou 等^[18]提供的颜色和谐模型认为,越接近的颜色组合,和谐度要越高.而对于可视化,首要目标还是增强不同类别之间的感知性差异,所以在计算颜色组合的分数时,感知性差异的权重应该更大.在本文的实验中取权重 $\alpha=10$.

3.2 优化限制

本文发现,除了以上述目标方程引导优化外,在实际解决问题时,还需要添加一些额外的限制要求,使最终的结果更加合理.

首先,在给定调色板的前提下,用户通常不希望调色板上颜色的色调发生巨大的变化,因此需要保持整体的色调分布.如图 2a 和图 2b 所示,在未添加该限制条件时,调色板中绝大多数初始的颜色经过优化后都发生了巨大的变化,例如,第 1 个的蓝色经过优化变成了绿色,因此需要固定每种颜色对应的色调,仅通过改变颜色的饱和度以及亮度进行优化.此外,为避免优化后的调色板中一些颜色的饱和度或亮度偏向极端的现象,如图 2c 所示,还需要对这 2 种颜色属性的优化范围进行限制.本文对颜色属性的值进行了归一化处理,将饱和度与亮度的优化范围设定为 $[0.35, 0.65]$. 添加 2 种优化限制后的结果如图 2d 所示.

图 2 不同条件下调色板的优化结果

3.3 优化实现

针对所提的需求,优化的实现主要是 2 个方面:

(1) 优化方法的选择与设计; (2) 颜色空间的选择.

首先,优化方法采用的是遗传算法,使用染色体表示不同变换方式下的调色板.根据设定,首先初始化生成包含 ω 个个体的种群,包括原始的调

色板,余下的则是在优化限制范围内采用均匀采样的方法随机生成.每个个体由多个三元数组组成,数组中的数值分别对应颜色的色调、饱和度以及亮度.在每次迭代中,使用目标方程对每个个体进行评价,选择 $v(v < \omega)$ 个最好的个体直接传递到下一代,同时从 ω 个个体中选取 $(\omega - v)$ 对个体用来产生剩余的下一代个体,目标方程越高的个体越容易被选择,将选中的每对个体进行线性组合得到下一代个体.同时,为了增加个体多样性并保证颜色的合理性,本文会随机选择染色体上某一数组,固定其色调值不变,在优化限制范围内对亮度及饱和度值进行随机替换.在本文方法中取参数 $\omega=20, v=5$.

在颜色空间的选择上,由于本文方法需要从色调、饱和度以及亮度 3 个方面进行颜色的变换,所以选用 HSL 颜色空间.但是,由于颜色和谐度模型和感知性差异 CIEDE2000 距离都是定义在 CIELAB 空间下的,所以还需要使用 CIELAB 颜色空间.因此,在本文方法中还定义了不同颜色空间相互转换的函数.

3.4 优化结果

本文对多个调色板进行了优化,部分结果如图 3 所示,每个调色板下面是对应的目标方程分数,本文将目标方程的分数同样进行了归一化处理.由于计算的是所有颜色组合之间的最小分数,所以目标方程分数普遍不高且不同调色板之间的差距不会特别大.

前面提到,最终结果受到调色板质量的影响,所以需要对调色板进行优化.经过对比,本文发现,当调色板的目标方程分数大于 0.2 时,就能保证最终可视化结果的质量,所以对于这些调色板可以不进行优化.可以看到,图 3 中优化后的调色板分数均大于 0.2.

图 3 本文调色板的优化结果

4 调色板的颜色分配

当给定调色板和多图表数据后, 调色板颜色分配就是找到数据类别与颜色之间的一种映射方式, 使不同类别的数据拥有不同的颜色. 定义 $A = \{A_1, A_2, \dots, A_m\}$ 表示具有 m 种类别数据的标签, $C = \{C_1, C_2, \dots, C_n\}$ 表示具有 n 种颜色的调色板, τ 表示一种颜色分配方案, 即 $\tau(A_i) = C_j, i \in [1, m], j \in [1, n]$.

默认按照数据给定顺序的颜色分配方案, 由于没有考虑到数据与最终可视化结果之间的几何对应关系, 导致最终结果感知性与和谐度较差, 因此为了提升最终可视化结果的质量, 就需要将相应的几何约束引入目标方程.

4.1 几何约束

虽然不同的可视化图表结构大有不同, 但是都可以由数据获得不同类别在最终结果中对应色块的位置和大小.

根据类别的输入顺序可以获得对应色块的位置关系, 如图 4 所示, 在 2 种布局下, 类别的输入顺序不同, 色块的邻接关系也不同, 在上面的布局中, A_1 相邻的色块是 A_2 和 A_5 ; 下面的布局中, A_1 相邻的色块是 A_2 和 A_4 . 定义 G_i 表示与 A_i 相邻色块的集合, 即 $G_i = [A_x, A_y, \dots], i \in [1, m]$.

对多图表的情况, 只需将不同图表中的邻接关系合并即可. 因此, 在图 4 中, $G_1 = [A_1, A_4, A_5]$.

图 4 色块位置的获取

根据数据的大小可以获得对应色块的大小, 如图 5 所示, 在不同图表中, 色块的大小也不同, 饼图中的 A_3 要明显小于柱状图中的 A_3 . 定义 ω_i 为 A_i 在不同图表中对应色块的大小, 数值越大, ω_i 也越大, 在图 5 中对应的圆也越大. 为了方便计算, 本文将 ω_i 进行了百分比化. 对于多图表的情况, 只需将不同图表中对应的 ω_i 求平均值即可, 再结合对应的邻接关系, 获得最终的几何约束.

图 5 色块大小的获取

4.2 目标方程

在进行颜色分配方案法优化之前, 首先要设计好目标方程, 为了使最终的结果拥有良好的感知性与和谐度, 目标方程设计为

$$E = \sum_{A_i \in A} E_H(A_i, G_i, \tau) + \beta_1 E_D(A_i, G_i, \tau) + \beta_2 E_L(A_i, \tau);$$

$$E_H(A_i, G_i, \tau) = \sum_{A_k \in G_i} (\omega_i + \omega_k) H(C_x, C_y);$$

$$E_D(A_i, G_i, \tau) = \sum_{A_k \in G_i} (\omega_i + \omega_k) D(C_x, C_y);$$

$$E_L(A_i, \tau) = \omega_i L(C_x, C_b);$$

$$\tau(A_i) = C_x, \tau(A_k) = C_y.$$

其中, E_H 用来衡量颜色分配方案的和谐度, 主要解决整体的和谐问题, 提高最终结果的美感效果; E_D 用来衡量颜色之间的感知性差异, 主要解决不同类别数据的感知性问题, 提高最终结果的感知性; E_L 用来衡量颜色与背景色之间的亮度差异, 也从一定程度上解决感知性的问题; C_b 表示背景颜色.

另外, 当观察一个目标物体时, 视觉系统对于它和背景的对比度具有很高的敏感性, 但对它的绝对颜色、质地和形状却不那么敏感. 如果没有足够的对比度, 那么对于物体的感知难度就会增加. 因此, 为了改善背景颜色对于可视化结果的影响, 引入了亮度差异, 通过计算 2 种颜色在 CIELAB 空间下亮度通道的 L_2 距离获得 $L(C_x, C_b)$ 的值.

感知性与和谐度的计算模型与调色板优化相同. 由于最终结果更侧重于不同类别数据的感知性差异, 所以在本文方法中取 $\beta_1 = 4, \beta_2 = 1$.

4.3 优化实现

对于颜色分配的优化, 同样采用遗传算法求解, 并且需要对几个关键步骤进行一定的调整.

首先是染色体编码, 采用的是实数编码, 使用不同的染色体代表不同的颜色分配方案, 由于颜

色分配案中, 每种类别的数据只能对应一种颜色, 所以染色体上不能有重复的基因。

其次是适应度函数, 为避免遗传算法的早熟问题, 本文采用了一种自适应的适应度函数^[19], 该方法会将迭代的次数引入适应度的计算当中, 避免了算法早期超级个体占据整个种群而导致的过早收敛问题。

交叉函数采用了常见的局部映射交叉方法, 该方法能够保证交叉后得到的染色体基因没有冲突, 从而保证结果中没有相同的颜色。

对于变异函数, 会出现 2 种不同的情况: (1) 当数据类别的数量与调色板的颜色数量一致时, 所有的颜色都会出现在颜色分配方案中, 只需要随机交换染色体上某 2 个位置的基因即可; (2) 当数据类别的数量小于调色板的颜色数量时, 调色板中会存在一些颜色未被选择, 为了增加整体的多样性, 随机选择染色体上的一个位置, 将该位置上的基因随机替换成未出现在该染色体上的颜色。

最后, 为了更快速的收敛, 设计良好的遗传算法参数也是非常必要的, 这些参数包括种群大小、交叉概率和变异概率等。但是, 当前研究并没有一个非常统一的规则帮助在不同的情况下设计良好的参数, 只能根据现有的经验或不断地尝试去调整。最终, 本文方法中将种群大小设定为 20, 交叉概率设定为 0.8, 变异概率设定为 0.05。

4.4 优化结果

本文使用不同类别数量的数据进行了颜色分配方案的优化, 分别使用了柱状图、饼图、堆叠柱状图以及堆叠图 4 种图表, 其中 9 类别的可视化结果如图 6 所示, 图 6a 为输入的调色板, 图 6b 为目标方程分数最小时的结果, 图 6c 为目标方程分数最大时的结果。可以明显看出, 无论从整体的视觉效果还是数据感知的难易程度, 图 6c 明显优于图 6b。此外, 优化结果与默认配色结果的对比如图 6c 与图 6d 所示, 其中图 6c 是本文的优化结果, 图 6d 是默认的配色结果, 从整体的视觉效果来看, 二者的区别并不大, 但是仔细观察每个图表, 不难发现本文优化结果的数据感知性要强于默认的配色结果。

图 6 是使用原本质量较高的调色板进行优化后的结果, 针对质量较低的调色板, 本文先对调色板进行了优化, 再对颜色分配方案进行优化, 即 2 种优化的异步形式, 结果如图 7 所示。图 7a 为输入的调色板, 图 7b 为优化后的调色板, 图 7c 为异步优化的结果, 与图 7d 所示的默认配色结果对比,

图 6 给定调色板的配色结果对比

异步优化的结果无论在整体和谐度还是数据感知的能力上都有了明显的提升。

此外, 本文方法还适用于同一图表多布局的情况, 如图 8 所示。图 8b 为本文的优化结果, 图 8c 为默认的颜色分配方案, 每个图表中, 左边是按类别顺序的布局, 右边是按数据大小顺序的布局。可以发现, 本文方法的优化结果无论在哪种布局下对于较小数据的感知能力都很好, 并且要优于默认的结果。

图7 低质量调色板的配色对比

图8 同一图表多布局的配色对比

5 用户调研

本文所提出的感知性与和谐度引导的多图表协同配色优化方法主要分为调色板优化和颜色分配优化2步,以下将通过用户调研的形式分别对原始调色板的颜色分配结果和调色板优化后的颜色分配结果进行评估。

5.1 原始调色板的颜色分配

为了验证本文目标方程设计的合理性以及可视化结果是否有很好的提升,主要从以下几个方面进行调研:(1)目标方程的设计是否具有合理性?能否对结果产生影响?(2)目标方程中是否真的需要颜色感知性与和谐度2种度量共同存在?(3)优化结果是否有相应的提升?

针对上述问题,本文进行了4组不同的用户调研,每组共5道题,并对结果进行分析,从而验证本文方法的有效性。

本文共招募了20名参与者,向他们展示了基于2种不同颜色分配方案,具有相同类别对应关系的多图表的可视化结果,要求他们在2种可视化结果中选取自己偏爱的一种。选择条件主要从感知性与颜色和谐度2个方面进行考虑,即首先选择更容易识别各图表类别信息的可视化结果,在2组结果类别区分能力差距不明显的情况下,选择他们认为整体颜色和谐度较高的一组。为了确保调研的有效性,本文在保证数据具有一定差异的前提下,随机生成了多组6~12类别数量的数据以供调研。

第1组实验是为了验证目标方程的合理性以及能否对可视化结果产生影响。将目标方程分数最高与最低的结果放在一起,让用户进行选择。实验结果如图9所示,其中二者皆可表示用户认为2种配色的效果整体差距不大。结果显示,根据本文目标方程优化的最优与最差结果之间有着明显的差异,有77%的用户选择了分数最高的结果。

图9 调研1第1组结果

第 2 组和第 3 组实验是为了验证目标方程中 2 种度量的必要性. 将优化结果分别与只包含一种度量的结果进行对比, 实验结果分别如图 10a 和 10b 所示. 根据结果可以发现仅有和谐度一种度量的结果明显较差, 分析和谐模型不难发现, 和谐模型更趋向于选择色调越接近、饱和度越相似的颜色, 这就导致最终的可视化结果在感知性上明显较弱. 根据图 10 结果可以看出, 单一度量下的优化质量较低, 大多数用户还是倾向于 2 种度量都具备的结果, 从而证明 2 种度量的必要性.

图 10 调研 1 第 2 组和第 3 组结果

第 4 组实验为了验证优化结果是否有提升, 将本文方法的优化结果与默认颜色分配方案下的结果放在一起, 让用户进行选择. 实验结果如图 11 所示, 结果表明有更多的用户喜欢本文方法的优化结果.

图 11 调研 1 第 4 组结果

5.2 调色板优化后的颜色分配

为了验证本文所提出的异步优化方法的有效性, 即先优化调色板再进行颜色分配优化. 本文还提出了同步优化方案, 即同时完成调色板和颜色分配方案的优化. 同样采用遗传算法实现同步优化的方式. 由于优化目标的改变, 需要对遗传算法中几个关键步骤进行重新的设计.

首先是染色体编码的设计, 根据所提出的要求可知, 最终的结果是在一个变化的调色板上选出一种较好的颜色分配方案. 所以, 在该问题中每个个体采用了双染色体的结构, 分别决定调色板的变化和颜色分配方案, 染色体 1 与调色板优化时的染色体一致, 染色体 2 与调色板颜色分配时一致.

目标方程的设计为

$$E = \gamma E_1 + E_2;$$

$$E_1 = \min(D[D(C_p, C_q)]), \forall p, q (p \neq q);$$

$$E_2 = \min \left[\sum_{A_k \in G_i} (\omega_i + \omega_k) \times H(C_x, C_y) + \omega_i \times L(C_x, C_b) \right],$$

$$i \in [1, m], \tau(A_i) = C_x, \tau(A_k) = C_y.$$

其中, E_1 由染色体 1 决定, 影响调色板的质量; E_2 由染色体 1 和 2 共同决定, 影响颜色分配方案的效果. 由于调色板优化计算的是颜色组合的最小分数, 为了统一, 颜色分配方案计算时, 也选择所有相邻色块颜色组合的最小分数. 本文方法中取 $\gamma = 4$.

交叉操作中, 染色体 1 和染色体 2 分别使用与之对应方法中的交叉操作, 之后产生的子代染色体进行相应的排列组合, 选择目标方程分数最高的组合作为最终的子代染色体. 变异操作与交叉相似, 也是对 2 条染色体分别进行变异操作后, 再重新组合. 其他设定与之前的基本一致.

针对上述 2 种优化方式, 本文采用同样的调研形式对以下几方面进行验证: (1) 调色板质量对可视化结果的影响? (2) 异步优化的方式是否对结果有提升? (3) 同步优化的方式是否具有良好的效果? (4) 异步优化是否优于同步优化?

第 1 组实验为了验证调色板质量对于可视化结果的影响, 将高质量调色板与低质量调色板的可视化结果放在一起, 让用户进行选择, 实验结果如图 12 所示, 有 81% 的用户选择了质量较高的调色板, 证明了调色板质量的影响.

第 2 组为了验证先优化调色板再优化颜色分配方案, 也就是 2 种优化异步的方式, 是否对可视化结果有提升, 将默认的配色结果与优化后的结

果放在一起,让用户进行选择,实验结果如图 13 所示,有 68%的用户选择了优化结果。

图 12 调研 2 第 1 组结果

图 13 调研 2 第 2 组结果

采用类似的形式验证优化与分配同时完成,也就是同步优化的方式,是否有效果,实验结果如图 14 所示,有 64%的用户选择了优化结果。综合上述 2 种情况,当所给的调色板质量较低时,对调色板进行优化是非常有必要的,也因此本文的方法能提升最终可视化结果的质量。

图 14 调研 2 第 3 组结果

最后,将异步优化与同步优化与 2 种方式进行对比,实验结果如图 15 所示,更多的用户还是倾向于异步优化的方式。

但这并不代表同步优化就无法得到更好的结果,如图 16 所示。图 16a 为输入的原始调色板,图 16b 为异步优化的结果,图 16c 为同步优化的结果,图 16d 为该组对比的用户调研结果,有超过一半的

用户选择了同步优化的结果。理论上同步优化更可能找到最优解,但是由于问题本身过于复杂,同步优化的搜索空间过大,导致无法找到相对于异步优化更优的组合解。

图 15 调研 2 第 4 组结果

a. 输入的调色板

b. 异步优化结果

c. 同步优化结果

d. 用户调研结果

图 16 异步优化与同步优化对比的调研结果

6 结 语

在本文的工作中,针对多图表协同配色的问题进行了 2 种情况的讨论.第 1 种是在给定一个质量较高的调色板后,对调色板的颜色分配方案进行优化,使最终得到的多个图表无论从感知性还是整体的和谐度上都能有很好的效果;第 2 种是在给定调色板的质量较低时,首先对调色板进行感知性与和谐度的优化,然后再对颜色分配方案进行优化,提高最终可视化结果的质量.本文方法同样适用于同一图表多布局协同配色的问题.

本文的创新点可以从以下几方面阐述.

(1) 本文研究的是多图表以及同一图表多布局的协同优化,目前关于此类问题的研究还较少.

(2) 本文从数据获取对应可视化结果中几何约束的方法具有很好的普适性,因此也可应用于单个图表的优化.

(3) 在调色板颜色分配问题上,本文方法对调色板颜色数量没有限制,且颜色选取属于全局选择.

(4) 在给定调色板质量较低的前提下,提出了异步和同步 2 种优化形式,提高可视化质量.

本文工作属于起步阶段,虽然经过用户调研,证明了方法的有效性,但依然有很多的地方需要进行优化与改进.

(1) 在几何约束的提取上,本文方法有些简单,没有考虑到坐标系尺度以及不用图表类型对于最终结果的影响;同时,单纯的将不同图表的临界关系进行合并,可能会对色块未相邻的图表产生影响,后续工作中需要完善几何约束提取的方法,如将色块在不同图表中相邻的次数也作为权重引入目标方程.

(2) 在目标方程的设计中,本文应用的 2 种度量存在重复的变量,因此会有一定的优化重叠,所以后续可以针对该问题,设计出一种新的标准,消除这种优化重叠现象.

本文所使用的优化方法均为遗传算法,虽然有遗传算法在求解近似最优解的问题上有很好的效果,但是在设计求解过程中发现,由于初始化种群的随机性,该算法求解具有一定的不稳定性.如图 17 所示,同样的调色板及数据,2 组优化结果(图 17b 和图 17c)有着明显差异.此外,算法的运行时间也比较长,特别是在调色板优化时,这种情况可能来源于遗传算法本身,也可能来源于本文方法

设计中还存在需要优化的地方,所以后续可以对该问题的优化方法进行更合理的设计.

图 17 不同初始化得到的不同优化结果

参考文献(References):

- [1] Trumbo B E. A theory for coloring bivariate statistical maps[J]. *The American Statistician*, 1981, 35(4): 220-226
- [2] Wang L J, Giesen J, McDonnell K T, et al. Color design for illustrative visualization[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2008, 14(6): 1739-1754
- [3] Harrower M, Brewer C A. ColorBrewer.org: an online tool for selecting colour schemes for maps[J]. *The Cartographic Journal*, 2003, 40(1): 27-37
- [4] Gramazio C C, Laidlaw D H, Schloss K B. Colorgical: creating discriminable and preferable color palettes for information visualization[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2017, 23(1): 521-530
- [5] Chuang J, Weiskopf D, Torsten M. Energy aware color sets[J]. *Computer Graphics Forum*, 2009, 28(2): 203-211
- [6] Lee S, Sips M, Seidel H P. Perceptually driven visibility optimization for categorical data visualization[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2013, 19(10): 1746-1757
- [7] Fang H, Walton S, Delahaye E, et al. Categorical colormap optimization with visualization case studies[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2017, 23(1): 871-880

- [8] Wang Y H, Chen X, Ge T, *et al.* Optimizing color assignment for perception of class separability in multiclass scatterplots[J]. IEEE Transactions on Visualization and Computer Graphics, 2019, 25(1): 820-829
- [9] Kim H R, Yoo M J, Kang H, *et al.* Perceptually-based color assignment[J]. Computer Graphics Forum, 2014, 33(7): 309-318
- [10] Lin S, Fortuna J, Kulkarni C, *et al.* Selecting semantically-resonant colors for data visualization[J]. Computer Graphics Forum, 2013, 32(3pt4): 401-410
- [11] Setlur V, Stone M C. A linguistic approach to categorical color assignment for data visualization[J]. IEEE Transactions on Visualization and Computer Graphics, 2016, 22(1): 698-707
- [12] Paterson I. A dictionary of colour: a lexicon of the language of colour[M]. London: Thorogood, 2004
- [13] Judd D B, Wyszecki G. Color in business, science, and industry[J]. Physics Today, 1953, 6(2): 18
- [14] Papachristos E, Tselios N, Avouris N. Bayesian modelling of colour's usage impact to Web credibility[C] //Proceedings of the 17th European Conference on Artificial Intelligence. Netherlands: Frontiers in Artificial Intelligence and Applications, 2006: 41-45
- [15] Ou L C, Luo M R. A colour harmony model for two-colour combinations[J]. Color Research & Application, 2006, 31(3): 191-204
- [16] Szabo F, Bodrogi P, Schanda J. Experimental modeling of colour harmony[J]. Color Research and Application, 2010, 35(1): 34-49
- [17] Lee Y K. Comparison of CIELAB Delta ΔE^* and CIEDE2000 color-differences after polymerization and thermocycling of resin composites[J]. Dental Materials, 2005, 21(7): 678-682
- [18] Ou L C, Yuan Y, Sato T, *et al.* Universal models of colour emotion and colour harmony[J]. Color Research and Application, 2018, 43(5): 736-748
- [19] Zhang Sicai, Zhang Fangxiao. A modified method to fitness function of genetic algorithms[J]. Computer Applications and Software, 2006, 23(2): 108-110(in Chinese)
(张思才, 张方晓. 一种遗传算法适应度函数的改进方法[J]. 计算机应用与软件, 2006, 23(2): 108-110)