

层次分析算法及优化在网络视频评比中的实践^①

李 凡^{1,2}, 杜义华¹

¹(中国科学院 计算机网络信息中心, 北京 100190)

²(中国科学院大学, 北京 100049)

通讯作者: 杜义华, E-mail: yhdu@cashq.ac.cn

摘 要: 随着新媒体时代的到来, 网络信息传播变得更加便捷, 很多单位开展了视频网上评比活动, 但往往对其缺少合理快捷的评估方法. 本文通过分析视频网上评比的影响因素, 构建出多指标的层次分析模型. 文章介绍了层次分析法及其基于 PageRank 的优化算法的主要步骤, 利用它求出各评比指标的权值, 从而给出评比计算方法. 本文的研究方法为处理多指标问题提供了一些参考价值.

关键词: 视频评比; 分析模型; 层次分析法; PageRank 算法; 权重值

引用格式: 李凡, 杜义华. 层次分析算法及优化在网络视频评比中的实践. 计算机系统应用, 2018, 27(7): 250-254. <http://www.c-s-a.org.cn/1003-3254/6426.html>

Practical Application of Analytic Hierarchy Process and Optimization in Online Video Evaluation

LI Fan^{1,2}, DU Yi-Hua¹

¹(Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China)

²(University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: With the coming of New Media Era, the spread of network information has become more extensive. Many online video evaluation activities were carried out by some institutions, while an evaluation method with rationality and fastness is lacked. In this study, we establish the hierarchial analysis model of many index signs on the basis of analyzing factors of online evaluation. The major steps of AHP and further PageRank optimization method are provided, which can be used to calculate weight values of index and obtain the final ratings. These research methods have certain reference values to dealing with multiple indexes appraisalment.

Key words: video rating; evaluation model; analytic hierarchy process; PageRank algorithm; weight value

手机的普及, 使网络信息的传播变得更加方便, 微视频、微博、微信等在新媒体时代下成为大量用户信息交互的平台. 合理利用网络渠道往往能提升活动的影响力. 近些年, 中科院及其附属单位开展了多次网上宣传评比活动, 诸如“镜头中的最美科学家”视频征集活动、科普短片征集评比活动等. 然而, 此类网上评比活动缺乏有效的评估方法.

对视频进行征集评比时, 一方面要考虑专家们的综合意见, 另一方面也要考虑网上网友们对视频播放

点赞转发等互动行为. 目前的多维指标综合评价模型主要使用用户自主赋予权重、均等分配权重等方法^[1], 存在很大的劣势, 诸如人为主观因素对权值分配干扰过强、指标太多时导致评价实施的复杂性等等. 网上评比需要考虑多维度的参考指标, 因此基于视频的多维交互量, 本文分析了评比影响要素并构建了层次分析模型, 引入层次分析法并加以算法改进, 确立了具体的指标权值, 给出了合理的评比方法.

① 收稿时间: 2017-11-07; 修改时间: 2017-11-27; 采用时间: 2017-12-07; csa 在线出版时间: 2018-06-27

1 视频网上评比模型

现有的视频网上评比方案存在许多问题. 用户投票评比法存在刷票、真正观看视频人数存在水分等数据失真问题. 仅依靠播放量等数据的评比法, 指标集比较粗糙, 数据不能够全面客观反映视频在观众中的传播影响力. 经过充分的问卷调查以及与专业人士的探讨, 得出如下结论: 视频网上评比, 不仅需要考虑直接的播放指标, 也需要考虑体现视频人气属性以及社交影响力的评论点赞以及转发等指标, 才能更全面科学地评测视频的传播效果及用户的喜好性. 利用层次分析法能很好地解决此问题. 因此本文主要从播放指标、评论点赞指标、转发指标这三个方面综合考虑视频评比问题.

各一级指标下的二级指标应客观真实且较易获取. 对于播放次数, 只有播放记录超过 2 分钟时才会后台记录为一次播放, 以减少视频点击后短期内退出等现象对指标的影响. 平均播放时长指标一定程度上反映出视频对用户的吸引力, 质量更好的视频往往会吸引用户观看更长时间. 用户对于偏喜好的视频, 倾向于留下表达自己想法的评论并点赞, 因此总评论数、总评论用户数、总点赞数指标能体现出一定偏好性. 转发功能能让视频更方便地在网络中传播, 用户对喜爱的视频可以随意转发, 因此总转发次数和总转发用户数可以反映视频的传播程度, 是一定程度上影响力的体现.

由上分析, 确立如表 1 的评比层次分析模型.

表 1 视频网上评比层次分析模型

视频综合评比 A	播放指标 B1	总播放次数 C11
		平均播放时长 C12
	评论点赞指标 B2	总评论数 C21
		总评论用户数 C22
		总点赞数 C23
	转发指标 B3	总转发次数 C31
总转发用户数 C32		

表 1 中左侧栏为目标层, 对应的是评比模型中视频的综合得分, 中间栏和右侧栏对应的是评比模型的准则层, C11、C12 是播放指标的二级指标, C21、C22、C23 是评论点赞指标的二级指标, C31、C32 是转发指标的二级指标.

选取样本集合为 S 需要对样本集中数据的每一个指标进行归一化处理. 归一化处理如公式 (1) 所示.

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

其中 x 为指标的原有值, x_{\max} 、 x_{\min} 分别为样本集中该指标的最大值和最小值. 归一化后的指标数值范围在 $[0, 1]$, 便于量化计算.

视频评比最终得分用公式 (2) 表示:

$$y = \sum_j x(j)w(j), j = 1, 2, \dots, 7 \quad (2)$$

式中 $x(j)$ 为样本集中某元素的第 j 个指标归一化后的数值, $w(j)$ 为第 j 个指标的权值, 通过层次分析法获取某指标对应的权值.

2 层次分析法

2.1 层次分析法权值计算

日常生活中人们经常遇到需要考虑多维多层次因素的决策问题, 由美国运筹学家托马斯·塞蒂提出的层次分析 (AHP) 能很好地解决此类问题, 它是一种定性和定量相结合的、系统化、层次化的分析方法^[2]. 一般会根据需要达到的最终目标剥离出问题的影响因素, 按照因素间的逻辑关系将各因素层次化划分, 形成一个层次分析模型. 之后逐步进行定性和定量分析, 确定从目标层到最低因素层的权值分配方案, 提供最终的决策方法.

层次分析法具体主要步骤如下:

1) 分析层次模型

分解出影响因素, 充分考虑因素间的内在关联, 将问题中遇到的各个因素进行逻辑分层, 建立起从最终决策到底层因素的多级层次. 在层次分析法中一般将层级划分为目标层、准则层、方案层等. 上层因素的权值等于其下属层各因素的权值之和.

2) 构造判断矩阵

对指标进行两两比较, 用 1 到 9 这 9 个数值表示指标间的相对重要程度^[3]. 每个数字级别代表的重要度含义如表 2 所示.

判断矩阵是否合理直接影响到最终结果的可信度. 由专家提供权威的判断, 对所有指标两两对比后, 得到该目标的判断矩阵 A.

$$A = \begin{bmatrix} \frac{w_1}{w_1} & \frac{w_1}{w_2} & \dots & \frac{w_1}{w_n} \\ \frac{w_2}{w_1} & \frac{w_2}{w_2} & \dots & \frac{w_2}{w_n} \\ \frac{w_3}{w_1} & \frac{w_3}{w_2} & \dots & \frac{w_3}{w_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{w_n}{w_1} & \frac{w_n}{w_2} & \dots & \frac{w_n}{w_n} \end{bmatrix} \quad (3)$$

表2 重要度比较表

级别	因素 <i>i</i> 和 <i>j</i> 间的关系
1	<i>i</i> 和 <i>j</i> 具有同等的重要性
3	<i>i</i> 和 <i>j</i> 稍微重要
5	<i>i</i> 比 <i>j</i> 明显重要
7	<i>i</i> 和 <i>j</i> 强烈重要
9	<i>i</i> 和 <i>j</i> 绝对重要
2、4、6、8	相邻级别的重要性中间值

3) 计算矩阵的特征值与特征向量

在层次分析法分析过程中, 矩阵的特征值、特征向量并非需要完全精确的数值, 只要求出近似的结果即可满足要求. 此处采用“方根法”近似计算判断矩阵的特征值与特征向量^[4,5].

求解过程如下:

① 求出判断矩阵每一行元素乘积的 *n* 次方根, 表示为:

$$\bar{w}_i = \sqrt[n]{\prod_{j=1}^n a_{ij}}, \quad i = 1, 2, \dots, n \quad (4)$$

② 对上述 \bar{w}_i 进行归一化处理, 即:

$$w_i = \frac{\bar{w}_i}{\sum_{i=1}^n \bar{w}_i} \quad (5)$$

得到 $W = (w_1, w_2, \dots, w_n)^T$ 即是矩阵的近似特征向量, 也是各评价指标被分配的向量权值.

③ 求出矩阵当前对应的最大特征值:

$$\lambda_m = \sum_i \left(\frac{(AW)_i}{w_i} \right) \cdot \frac{1}{n} \quad (6)$$

4) 一致性检验

在判断矩阵 *A* 中, 满足如下关系:

$$a_{ij} = a_{ik} \times a_{kj}, \quad i, j, k = 1, 2, \dots, n \quad (7)$$

一致性检验中需要用到一致性指标 *CI*, 其表示当前判断矩阵不一致性的程度, *CI* 的值与最大特征向量 λ_m 以及判断矩阵的阶数 *n* 有关, *CI* 值越小的判断矩阵越趋向于完全一致性^[6,7]. 计算 *CI* 的方法如公式 (8)

所示:

$$CI = \frac{\lambda_m - n}{n - 1} \quad (8)$$

其中, *n* 为判断矩阵的阶数, λ_m 为当前判断矩阵的最大特征值.

计算一致性比例 *CR*:

$$CR = \frac{CI}{RI} \quad (9)$$

其中, *RI* 为平均随机一致性指标, 查阅相关资料, 得到常用的 *RI* 数据如表 3 所示.

表3 一致性指标

阶数	1	2	3	4	5	6	7	8
<i>RI</i>	0	0	0.52	0.89	1.12	1.24	1.32	1.41

所有阶数小于等于 2 的判断矩阵都满足完全的一致性. 此外, 当 $CR < 0.1$ 时, 矩阵满足一致性要求, 否则需对矩阵元素进行调整来重新计算矩阵的特征值与特征向量, 直到该判断矩阵达到最终的一致性要求.

2.2 基于 PageRank 的 AHP 优化算法

上述层次分析法, 依靠判断矩阵实现了指标的权值分配. 但是, 它仅仅反映了各指标间的独立比较关系, 并没有考虑到各指标间的依赖关系. 因此, 本文在对指标分析权重时, 考虑指标间的依赖关系, 参考谷歌 PageRank 算法, 对原始 AHP 算法加以优化.

1) PageRank 算法介绍

PageRank 算法是一种应用广泛的链接分析算法, 由 Larry Page 和 Sergey Brin 在研究网页排序算法时提出的, 是一种评估网页重要性的测量方法^[8].

其基本思想是: 访问越多的网页, 其网页质量偏向于越高. 某一页面的 PageRank 值是通过递归算法计算所有链入它的页面的因子的和得到的^[9]. 因而一个网页要想有较高的等级, 必须具有较多的链入页面, 高质量的链入页面越多, 会导致等级越高^[10].

2) 改进 AHP 指标权值

通过普通层次分析法得到的指标权值分配, 初步体现了指标间层次因果关系, 但是还存在一些局限性, 比如没有考虑到同级指标间的依赖关系. 因此参考 PageRank 算法对已得到的权值向量加以改进.

对于某个指标 *A1*, 如果它依赖于指标 *A2*, 则认为 *A1* 需要将自己的部分权值因子分配给 *A2*. 诸如视频的播放量指标就依赖于转发数、点赞数等指标, 所以需要播放量指标的权值分一部分给转发数、点赞数

等指标. 这和 PageRank 算法的链接传递关系很相似.

对于某一指标 i , 将它的被依赖指标集记为 S , 考虑依赖关系得到:

$$W_{P_i}' = W_{P_i} + \sum_{P_j \in S} \frac{W_{P_j}}{N(j)} \quad (10)$$

其中, W_{P_i}' 是更新后的权值, $N(j)$ 是指标 P_j 对应的依赖指标数目.

每一轮更新后, 指标权值就会发生相应变化. 需要对更新后的权值作归一化处理:

$$W_{P_i} = W_{P_i}' / \sum_{j \in A} W_{P_j}' \quad (11)$$

重复步骤 (10) 和 (11) 作迭代运算, 直到数值趋于稳定.

矩阵变更运算是否已经收敛的判定条件是:

$$\delta = \frac{\sum_{i=1}^{n \times n} \left| \frac{W_{P_i}' - W_{P_i}}{W_{P_i}} \right|}{n \times n} \quad (12)$$

当 $\delta \leq 0.1$ 时, 可以认为矩阵的变更已趋于稳定, 可以停止迭代运算.

经过 PageRank 算法思想改进的指标权值, 体现了各指标间的内在关联, 符合整体性考虑的要求.

3 AHP 权值计算

1) 构造层次分析模型

构建出如表 1 的层次分析模型.

2) A-B 层间的权向量

首先需要建立合理的判断矩阵, 它是指标元素两两重要性比较的直观数据体现. 根据反映指标相对重要性的九级度量法, 初步构造出目标层与一级指标层之间的判断矩阵如表 4 所示.

表 4 判断矩阵 A-B

A	B1	B2	B3
B1	1	1	5
B2	1	1	3
B3	1/5	1/3	1

下面计算矩阵特征向量与特征值.

根据公式 (4) 可以得到:

$$\overline{w}_1 = \sqrt[3]{1 \times 1 \times 5} = 1.7100, \overline{w}_2 = \sqrt[3]{1 \times 1 \times 3} = 1.4422,$$

$$\overline{w}_3 = \sqrt[3]{1/5 \times 1/3 \times 1} = 0.4055$$

根据公式 (5) 可以得到:

$$w_1 = \frac{\overline{w}_1}{\sum_{i=1}^n \overline{w}_i} = \frac{1.7100}{1.7100 + 1.4422 + 0.4055} = 0.4806$$

同理, $w_2 = 0.4054, w_3 = 0.1140$.

即 A-B 的特征向量为:

$$W = [0.4806, 0.4054, 0.1140]^T$$

由公式 (6) 可以得到:

$$AW = \begin{bmatrix} 1 & 1 & 5 \\ 1 & 1 & 3 \\ 1/5 & 1/3 & 1 \end{bmatrix} \begin{bmatrix} 0.4806 \\ 0.4054 \\ 0.1140 \end{bmatrix} = \begin{bmatrix} 1.4560 \\ 1.2280 \\ 0.3453 \end{bmatrix}$$

$$\lambda_m = \sum_i \left(\frac{(AW)_i}{w_i} \right) \cdot \frac{1}{n}$$

$$= \frac{1.4560}{3 \times 0.4806} + \frac{1.2280}{3 \times 0.4054} + \frac{0.3453}{3 \times 0.1140} = 3.0292$$

下面检验该判断矩阵是否满足一致性要求. 由公式 (8)、(9) 可以得到:

$$CI = \frac{\lambda_m - n}{n - 1} = \frac{3.0292 - 3}{3 - 1} = 0.0146$$

$$CR = \frac{CI}{RI} = \frac{0.0146}{0.53} = 0.0275$$

因为 $CR < 0.1$, 故矩阵满足一致性要求.

3) B-C 层间的权向量

参照 2) 所示步骤, 分别求出以下几个 B-C 层次间特征向量为:

$$W_{B1-C1} = \begin{bmatrix} 0.8333 \\ 0.1667 \end{bmatrix}, W_{B2-C2} = \begin{bmatrix} 0.6480 \\ 0.1220 \\ 0.2297 \end{bmatrix}$$

$$W_{B3-C3} = \begin{bmatrix} 0.7500 \\ 0.2500 \end{bmatrix}$$

4) 计算综合权重表

根据以上步骤, 利用权向量对权值的分配, 得到综合权重表如表 5 所示.

表 5 综合权重表

目标层	一级指标	一级指标 权值分配	二级指标	二级指标 权值分配	二级指标 具体权值
A	B1	0.4806	C11	0.8333	0.4005
			C12	0.1667	0.0801
	B2	0.4054	C21	0.6480	0.2627
			C22	0.1220	0.0495
	B3	0.1140	C23	0.2297	0.0931
			C31	0.7500	0.0855
			C32	0.2500	0.0285

5) 基于 PageRank 的优化算法对权重表修正

考虑各指标间的依赖与被依赖关系, 得到指标依赖关系图如图 1 所示。

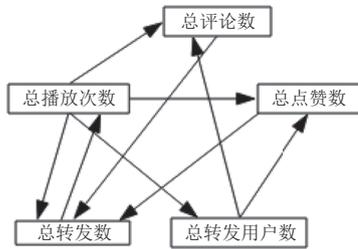


图 1 指标依赖关系图

图 1 中总播放次数 C11 指标依赖于总评论数 C21、总点赞数 C23、总转发次数 C31、总转发用户数 C32, 同时 C11 也被总转发数 C31 指标所依赖。其他指标也都存在各自的依赖指标集以及被依赖指标集。

根据公式 (10)、(11)、(12) 作多次迭代运算, 直到各指标权值实现稳定收敛。

经过基于 PageRank 的优化算法, 得到更新后的权值表如表 6 所示。

表 6 AHP 优化后的综合权重表

目标层	一级指标	一级指标 权值分配	二级指标	二级指标 权值分配	二级指标 具体权值
A	B1	0.3521	C11	0.8333	0.3208
			C12	0.1667	0.0313
			C21	0.6480	0.1332
	B2	0.2795	C22	0.1220	0.0370
			C23	0.2297	0.1093
			C31	0.7500	0.2969
B3	0.3384	C32	0.2500	0.0715	

4 实验分析

为验证本文提出的评比模型的有效性, 进行了模拟实验。从数据集中选取了十个征选视频, 首先由相关工作人员进行打分, 得出专家评估分。再分别用本文的层次分析法模型以及均等权重法进行量化打分, 计算出各自的综合评比分数。由实验结果表 7 看出, 本文评比模型计算出的评分数据与工作人员给出的评分数据更为接近, 说明其具有一定合理性。

5 结语

本文提出了基于层次分析优化算法的视频网上评比方法, 首先分析了评比因素并建立了层次分析模型,

通过层次分析优化算法得出各指标的被赋予权重, 最后计算出在该模型下的综合得分, 经实验验证该方法具有良好的分析效果。与传统的人为主观赋予权重方法相比, 减少了主观因素的不利影响, 合理有效地解决了多指标应用场景的综合评比问题。本文引入的方法简单、易于操作, 能很好地解决视频网上评比问题, 为新媒体时代其他传播工作的研究也提供了一些参考价值。

表 7 实验结果

编号	专家评分	本文模型	均等权重法
1	0.70	0.6828	0.7623
2	0.79	0.7850	0.7142
3	0.89	0.9103	0.8130
4	0.73	0.7376	0.6476
5	0.85	0.8670	0.9033
6	0.66	0.6406	0.7118
7	0.63	0.6098	0.5100
8	0.55	0.5697	0.6031
9	0.65	0.6381	0.6077
10	0.73	0.6730	0.7255

参考文献

- 杨艳屏. 基于层次化分析的全网业务健康度评价. 计算机系统应用, 2013, 22(5): 9-13.
- 屈正庚. 层次分析法在商洛市农家乐评价体系中的研究. 计算机系统应用, 2016, 25(9): 236-240. [doi: 10.15888/j.cnki.csa.005378]
- 沈惠文. 层次分析法在项目风险评价中的应用. 自动化与仪器仪表, 2016, (8): 118-120.
- 李鑫, 李京春, 郑雪峰, 等. 一种基于层次分析法的信息系统漏洞量化评估方法. 计算机科学, 2012, 39(7): 58-63.
- 王志, 刘艳辉, 杨欢. 层次分析法在软件过程度量中的应用. 计算机工程与设计, 2017, 38(1): 144-148.
- 杨秀梅, 孙咏, 王美吉, 等. 层次分析法在网络用户信任度评价模型中的研究. 计算机系统应用, 2016, 25(3): 267-270.
- 谢海涛, 仲梁维. 基于层次分析法的装配序列评价技术. 计算机系统应用, 2012, 21(2): 72-76, 149.
- 李稚樞, 杨武, 谢治军. PageRank 算法研究综述. 计算机科学, 2011, 38(S1): 185-188.
- 高春生, 吴霖. 基于 Google PageRank 算法的围棋棋手排名方法. 信息技术, 2016, 40(11): 42-45, 51. [doi: 10.3969/j.issn.1672-9528.2016.11.008]
- 王冲, 纪仙慧. 基于用户兴趣与主题相关的 PageRank 算法改进研究. 计算机科学, 2016, 43(3): 275-278, 312. [doi: 10.3969/j.issn.1672-9528.2016.11.008]

10.11896/j.issn.1002-137X.2016.03.051]

www.c-s-a.org.cn

www.c-s-a.org.cn