

一种等价类区分的多属性评价指标约简算法

朱海萍¹, 王云岚², 李增智¹

(1. 西安交通大学计算机系统结构与网络研究所, 710049, 西安;
2. 西北工业大学高性能计算研究与发展中心, 710072, 西安)

摘要: 针对多属性评价中消除评价指标之间相关性的问题,从指标的组成出发,将评价指标分为元指标和合成指标,并分别给出了线性合成指标和非线性合成指标的相关度计算方法. 借鉴 Dijkstra 算法,提出了基于指标等价类的合成指标动态约简算法,即通过等价类划分来缩小参与约简的指标范围,并充分利用前一次约简以求得最终指标约简的结果. 仿真实验表明,随着指标数目的增多,所提算法比直接使用改进 Dijkstra 算法进行指标约简的运行时间要少,可提高大规模指标约简的执行效率.

关键词: 多属性评价;约简算法;相关度;指标等价类

中图分类号: TP393 **文献标志码:** A **文章编号:** 0253-987X(2009)04-0026-04

Eliminating the Correlation of Multi-Attribute Evaluation Based on Equivalence Class Partition

ZHU Haiping¹, WANG Yunlan², LI Zengzhi¹

(1. Institute of Computer Architecture and Networks, Xi'an Jiaotong University, Xi'an 710049, China; 2. Center for High Performance Computing, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: To eliminate the correlation among evaluation indexes, the evaluation indexes are divided into element indexes and synthetic indexes. Then the methods of calculating the correlation degree of linear synthetic indexes and nonlinear synthetic indexes are given. An algorithm of eliminating the correlation of synthetic indexes based on index equivalence class is proposed in terms of Dijkstra algorithm. The reduction ranges of evaluation indexes can be reduced, and the anterior reduction results are used to increase the efficiency. Comparisons with the modified Dijkstra index reduction algorithm show that the proposed algorithm based on equivalence class spends less running time. The more indexes there exist, the more efficient the proposed algorithm is.

Keywords: multi-attribute evaluation; reduction algorithm; correlation degree; index equivalence class

多指标综合评价是对多属性体系结构描述的对象系统做出全局性、整体性的评价,亦称多属性综合评价^[1]. 目前的研究大致可以分为两类,一类是对评价指标体系的研究,另一类是对综合评价方法的研究. 在现实世界中,事物的复杂性和评价目的多样性,决定了评价指标集的复杂性和多变性,所以评价指标的选取应该遵循整体性原则、可比性原则、均匀性原则^[2]等. 片面地追求指标体系的全面性,容易造成评价指标的冗余,同时会提升评价过程的复杂度. 已有的相关性指标约简,无论是定量或是定性计算,大都是针对指标集确立的^[3-6],很少考虑到评价任务、评价目标或是因评价主体的改变引起的评价指标

收稿日期: 2008-08-20. 作者简介: 朱海萍(1974—),女,在职博士生,讲师;李增智(联系人),男,教授,博士生导师.
基金项目: 国家自然科学基金资助项目(90304006,60673170).

标的重新选取问题,以及由此带来的评价指标相关性约简的动态更新问题.

另一方面,已有的评价指标体系也没有对指标进行细化区分.实际上,在各项评价指标中除了单个指标,还有很多反映指标关联的综合指标(合成指标),它们是在单个指标的基础上形成的具有线性或非线性函数关系的指标.评价指标约简就是如何动态地消除合成指标之间的相关性.

本文从研究指标的组合格形式出发,把评价指标分为元指标和合成指标,给出了合成指标和指标相关度的定义,并借鉴 Dijkstra 最短路径算法^[7],提出了一种基于指标等价类的动态指标约简算法.

1 多属性评价指标约简

1.1 评价指标分类

网络管理中的性能表达式是典型的合成指标^[8].合成指标的约简就是要将多个具有相似目的和描述的合成指标简化成一组新的相互无关或相关性很小的指标.

定义 1 在多属性评价中,将反映评价指标之间关联性(或冗余性)的参数定义为指标相关度.

定义 2 将相对独立的评价指标,如带宽、时延、比特率和丢包率等定义为元指标.由若干个元指标组成的集合 $P_1=(e_1,e_2,\cdots,e_n)$ 定义为元指标集.

定义 3 将一组元指标线性组合而成的指标定义为线性合成指标 $\mu_j=\sum_{i=1,\cdots,n}\omega_ie_i$,其中 $e_i\in P_1,1\leq i\leq n,1\leq j\leq k,\omega_i$ 是 e_i 的权重. ω_i 越大,意味着元指标 e_i 取值的变化对合成指标 μ_j 取值的影响越大.

定义 4 将一组元指标非线性组合而成的指标定义为非线性合成指标 $\eta_j=f_j(e_1,e_2,\cdots,e_n)$,其中 $e_i\in P_1,1\leq i\leq n,1\leq j\leq K$.

线性合成指标和非线性合成指标共同组成合成指标集合 $P_2=(\eta_1,\eta_2,\cdots,\eta_K,\mu_1,\mu_2,\cdots,\mu_k)$.评价指标集合 $P=P_1+P_2$ 且 $P_1\cap P_2=\varnothing$.

定义 5 建立在相同元指标集合 $(e_1,\cdots,e_j)$ 上的一组线性或非线性合成指标构成一个合成指标等价类 $E,E(e_i,\cdots,e_j)=(\gamma_1,\gamma_2,\cdots,\gamma_i)$,满足 $(e_i,\cdots,e_j)\subset P_1,\gamma_i\in P_2,E\subset P_2$.

等价类描述了这样一组线性或非线性合成指标,即任何 $(e_i,\cdots,e_j)$ 之外的元指标的变化不会影响合成指标的值.

1.2 线性合成指标的相关度

$\forall e_i\in P_1(1\leq i\leq n),\forall \mu_1,\mu_2\in P_2,\mu_1=$

$\sum_{i=1,\cdots,n}\omega_ie_i,\mu_2=\sum_{i=1,\cdots,n}\omega'_ie_i$,则 μ_1 和 μ_2 的相关度与 2 个元指标的权值差相关,即

$$W_{12}^\mu=\sum_{i=1,\cdots,n}|\omega_i-\omega'_i| \tag{1}$$

$\forall \mu_j\in P_2,\forall e_i\in P_1(1\leq i\leq n,1\leq j\leq k)$,则所有线性合成指标的相关度可以表示为一个对称矩阵

$$W(\mu)=\begin{bmatrix} 0 & W_{12}^\mu & \cdots & W_{1k}^\mu \\ W_{21}^\mu & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & W_{(k-1)k}^\mu \\ W_{k1}^\mu & \cdots & W_{k(k-1)}^\mu & 0 \end{bmatrix}$$

其满足 $W_{ij}^\mu=W_{ji}^\mu$.

1.3 非线性合成指标的相关度

对于非线性合成指标,通过判断函数增减特性来反映非线性合成指标之间的相关性.构造一个向量 $(\omega_{j1},\omega_{j2},\cdots,\omega_{jn}),1\leq j\leq K$,来描述非线性合成指标 $\eta_j=f_j(e_1,e_2,\cdots,e_n)$ 和元指标 $(e_1,e_2,\cdots,e_n)$ 之间的相关性,其中 $\omega_{ji}(1\leq i\leq n)$ 是根据表 1 确定的 η_j 在 e_i 上的赋值. $\forall \eta_1,\eta_2\in P_2,\eta_1=f_1(e_1,e_2,\cdots,e_n),\eta_2=f_2(e_1,e_2,\cdots,e_n)$.对于 f_1 和 f_2 ,通过取值空间切分,总能确定在某一取值空间上 f_1 或 f_2 是某个元指标变量 $e_i(1\leq i\leq n)$ 的递增或递减函数.

表 1 非线性合成指标赋值表

ω_{ji}	含义
1	f_j 是 e_i 的增函数
-1	f_j 是 e_i 的减函数
0	f_j 与 e_i 不相关

综上,定义 2 个非线性合成指标的相关度

$$W_{ij}^\eta=\sum_{p=1}^n|\omega_{ip}-\omega_{jp}| \quad 1\leq i,j\leq K \tag{2}$$

W_{ij}^η 越大,表示 2 个非线性合成指标的相关度越小;反之,则表示 2 个非线性合成指标的相关度越大.可以理解为,若 2 个非线性合成指标在某个取值区间上对大多数的元指标同为增函数或者同为减函数,认为其相关性较小,否则认为相关性较大.因此,所有非线性合成指标的相关度仍然可以表示为一个对称矩阵

$$W(\eta)=\begin{bmatrix} 0 & W_{12}^\eta & \cdots & W_{1K}^\eta \\ W_{21}^\eta & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & W_{(K-1)K}^\eta \\ W_{K1}^\eta & \cdots & W_{K(K-1)}^\eta & 0 \end{bmatrix}$$

其满足 $W_{ij}^\eta=W_{ji}^\eta$.

2 基于指标等价类的动态指标约简算法

2.1 问题的提出

给定连通图 $G(V, E)$, 其中 G 是由一个非空有限集合 $V(G)$ 与其中某些元素的无序对集合 $E(G)$ 构成的二元组. 每个节点代表一个具体的评价指标, 节点之间连线的权值反映 2 个指标的相关度. 那么, 指标约简就是在线性或非线性合成指标所构成的集合中选取相关度最小的一组线性指标 $(\mu_i, \dots, \mu_j)$ 或非线性指标 $(\eta_i, \dots, \eta_j)$, 而对应最短路径算法就是在图 G 中找到任意 2 个合成指标之间的一条最短路径. 下面若不做特殊说明, 参与约简的指标代表线性或非线性合成指标.

2.2 算法描述

借鉴 Dijkstra 算法, 本文提出了基于指标等价类的合成指标动态约简算法. 其主要思想是, 在进行指标约简之前把合成指标划分为若干个指标等价类, 对于新增或删除一个合成指标而导致的指标变化, 可通过判断确定它是否属于一个已知的等价类, 从而把最短路径的动态更新, 即基于 Dijkstra 的指标约简, 限定在一个等价类的内部进行, 以避免全面重新计算最短路径.

基于等价类的动态指标约简算法描述如下.
输入: 一组原始的合成指标
$$P' = (b_1, b_2, \dots, b_i) \quad P' \subset P_2$$

条件: 增加、删除一个合成指标 P^* , 或改变兴趣度 ∂ 等参数.

输出: 一组合成指标的动态约简结果 $(p_1, p_2, \dots, p_k)$.

步骤 1: 根据定义 5, 将一组期待约简的指标集 P' 划分为 x 个等价类.

步骤 2: 在每个等价类的内部使用改进 Dijkstra 算法进行指标约简, 得到 x 个等价类的指标约简结果 $(A_1, A_2, \dots, A_x)$, $A_i = (a_i^1, a_i^2, \dots, a_i^{\mu_i})$, μ_i 是约简后 A_i 中的指标个数, $1 \leq i \leq x$.

步骤 3: 使用改进 Dijkstra 算法在约简后的等价类之间进行指标约简, 从而完成了一次完整的指标约简, 得到一组指标约简结果 $(p_1, p_2, \dots, p_k) \subset (A_1 \cup A_2 \cup \dots \cup A_x)$, 转至步骤 5.

步骤 4: 对于由指标数或者兴趣度等参数改变引起的动态指标约简, 根据输入条件的变化, 执行图 1 操作.

步骤 5: 算法结束, 得到指标约简后的一组线性或非线性合成指标.

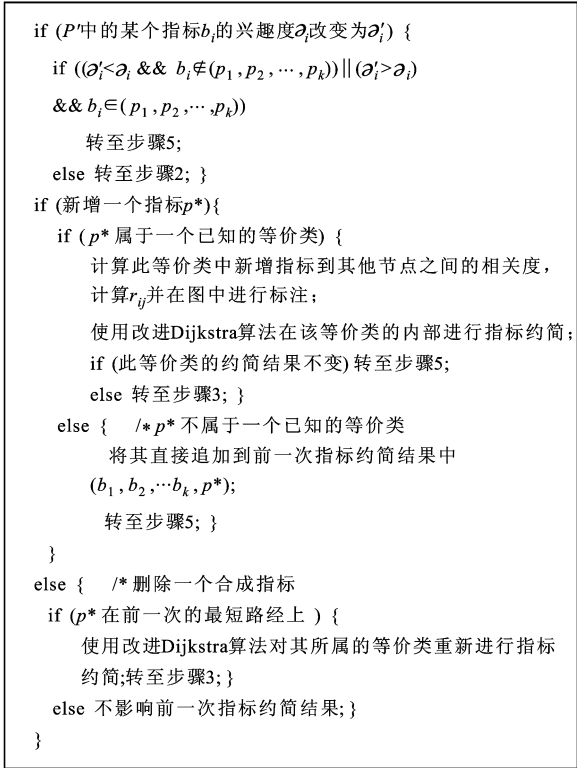


图 1 依据输入条件变化执行的操作

改进 Dijkstra 算法主要体现在每次需要扫描的点(指标)按其所在边的权值(指标之间的相关度)进行顺序排列, 这样每循环一次即可取得符合条件的点, 在一定程度上提高了算法的执行效率.

3 仿真实验及分析

传统 Dijkstra 算法的时间复杂度为 $O(n^2)$, 相比而言, 本文的指标约简算法通过指标分类, 将指标约简限定于一个等价类的内部, 这样在一定程度上降低了指标约简的总计算次数^[9]. 为了测试本文算法的运行效率, 在 2.0 GHz, 512 MB 的机器上, 开发平台是 Visual C++6.0, 并采用标准 C 实现了改进的 Dijkstra 和基于指标等价类的指标约简算法.

实验中, 随机产生 n 个线性合成指标 $\mu_j = \sum_{i=1, \dots, 10} \omega_i e_i$, 元指标数为 10. 表 2 显示了 $n=16, 32, 64, 128, 1\ 024$ 且动态增加一个指标 μ^* 的情况下, 程序重复运行 100 次之后本文约简算法和 Dijkstra 算法进行指标约简的平均计算次数比较. 可以看出, 前者的平均计算次数较之 Dijkstra 算法有所下降, 随着节点数、等价类数增多, 下降趋势越明显.

$n=100, 200, 300, 400, 500$, 程序重复运行 100 次之后两种算法的平均运行时间如图 2、图 3 所示.

表 2 两种算法的平均计算次数比较

节点数	平均计算次数	
	Dijkstra 算法	本文算法
16	120	108
32	496	451
64	2 016	1 749
128	8 128	6 623
1 024	523 776	476 457

可以看出,相比 Dijkstra 算法,随着指标数的增多,本文算法所需的运行时间更少。

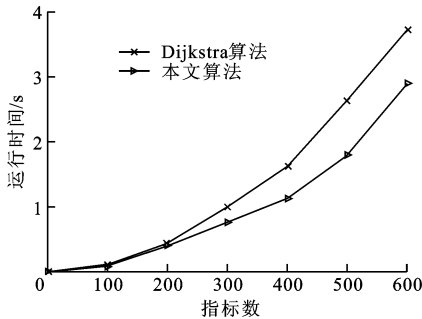


图 2 指标数对运行时间的影响

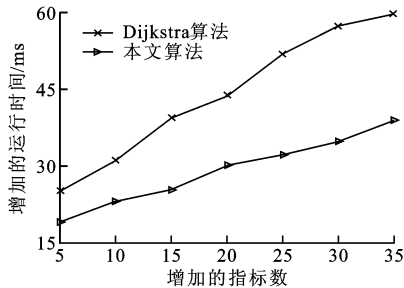


图 3 增加的指标数对运行时间的影响

4 结束语

在现有的质量评价研究中重点解决的是,评价体系、评价流程和针对特定领域特定业务的评价问题,而评价指标的选取一般是事先给定的,带有一定的经验性,不能动态更新,也不可约简。为此,本文提出了一种基于指标等价类划分的动态指标约简算法,并给出了算法运行时间分析。下一步可以考虑实际中该评价指标约简的优化问题,并研究如何构造反映指标之间逻辑关联的多属性综合评价算法。

参考文献:

[1] CHEN S J,HWANGC L. Fuzzy multiple attribute decision making: methods and application [M]. New York, USA: Springer,1992; 465-486.

[2] 郭亚军. 综合评价理论、方法及应用[M]. 北京:科学出版社,2007.

[3] 李辉,张爽. 全面质量管理的电子服务质量评价体系[J]. 西安交通大学学报(社会科学版),2008(2): 33-37.

LI Hui,ZHANG Shuang. A study on the evaluation of E-service quality based on total quality management [J]. Journal of Xi'an Jiaotong University(Social Sciences), 2008(2):33-37.

[4] RAIKO T, ILIN A, KARHUNEN J. Principal component analysis for large scale problems with lots of missing values[C]//18th European Conference on Machine Learning. Berlin, Germany: Springer-Varlag, 2007: 691-698.

[5] 张大陆,杨哲,姚进. 分布式电子商务中服务评价指标相关性消除方法[J]. 同济大学学报,2006, 34(3): 401-404.

YANG Dalu, YANG Zhe, YAO Jin. Research on elimination the correlation of service evaluating index in distribution e-commerce[J]. Journal of Tongji University, 2006, 34(3): 401-404.

[6] 胡周君,胡志刚,李林. 一种基于性能评估的元任务调度算法[J]. 西安交通大学学报, 2008, 42(8): 972-976.

HU Zhoujun, HU Zhigang, LI Lin. Performance evaluation based meta-task scheduling algorithm[J]. Journal of Xi'an Jiaotong University, 2008, 42(8): 972-976.

[7] GARCIA N M, LENKIEWICZ P, FREIRE M M, et al. On the performance of shortest path routing algorithms for modeling and simulation of static source routed networks: an extension to the Dijkstra algorithm[C]// The 2nd International Conference on Systems and Networks Communications. Piscataway, NJ, USA: IEEE, 2007; 60-65.

[8] 岑贤道. 网络管理协议及应用开发[M]. 北京:清华大学出版社,1999.

[9] AHO A V, HOPCROFT J E, ULLMAN J D. Data structures and algorithms [M]. Hong Kong, China: Pearson Education Asia Limited,2007.

(编辑 苗凌)