

结合聚类分析与信息熵赋权的构件选择方法*

田 帅 汪海涛 姜 瑛 陈 星

(昆明理工大学信息工程与自动化学院 昆明 650500)

摘 要 为解决多位专家从构件库系统检索结果中选出最适合用户需求的构件这一问题,论文提出了一种结合聚类分析与信息熵赋权的构件选择方法。首先由业内专家对系统自动检索出的构件进行再次评估选择,然后通过论文提出的构件选择方法给出一个综合的构件选择结果,把系统自动检索与人工经验选择结合起来,多角度全面地评估选择构件,以利于得到最优和最适合用户需求的构件。该方法不但结合了专家们对用户需求及构件的整体把握,更通过科学有效的方法计算出最优构件,选出的构件能更好地发挥软件复用的优势。并给出了一个实例,通过实验对该构件选择方法进行了分析说明,为进一步的研究工作奠定了基础。

关键词 构件选择; 最优选择; 聚类分析; 信息熵; 赋权

中图分类号 TP18 **DOI:**10. 3969/j. issn. 1672-9722. 2017. 12. 024

A New Software Component Selection Method
Based on Clustering and Information Entropy Weighting

TIAN Shuai WANG Haitao JIANG Ying CHEN Xing

(Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500)

Abstract In order to solve the problem of selecting the most suitable components which meet the needs of users from the search results of component library system by experts, this paper proposes a component selection method based on clustering and information entropy weighting. Firstly, industry experts evaluate and select components again from which automatically retrieved by the system, then the component selection method which proposed in this paper gives a comprehensive component selection result, combining the system automatic retrievals with artificial experience choice, comprehensively evaluating and selecting components from multi-angle so as to select the most optimal and appropriate components to meet the users' needs. This method not only combines the overall grasp of the experts on the needs of users and components, but also computes and selects the optimal components through scientific and effective method, so that the selected components can better play to the advantages of software reuse. And this paper also gives an example to analyze and explain the components selection method by the way of experiments, which lays the foundation for the further research. Research work in this paper has some theoretical research value and practical significance.

Key Words component selection, optimal choice, clustering analys, information entropy, weighting

Class Number TP18

1 引言

1968 年,McIlroy 在北大西洋公约组织(NATO)会议上提交的《大量生产的软件构件》一文中,首次提出了软件复用和构件库概念。至此构件开发思想便应运而生。随着时间的推移,涌现出大量的构件评估方式和方法,可复用构件的规模也不断扩

大,由此构件的分类检索便成为一个亟待解决的问题。基于此,国内外专家结合现有的知识理论,构建了各种分类及检索模型^[1],如基于剖面分类的检索、基于关键词的分类检索等等,随着互联网大数据时代的到来,数据信息量急剧增长,在这个大环境背景下构件的开发及复用也获得了前所未有的发展,并且构件开发技术愈发纯熟,可信度明显提

* 收稿日期:2017 年 6 月 6 日,修回日期:2017 年 7 月 7 日
基金项目:国家自然科学基金项目(编号:61462049)资助。
作者简介:田帅,男,硕士研究生,研究方向:软件工程。汪海涛,女,副教授,硕士生导师,研究方向:软件工程。姜瑛,女,博士,教授,硕士生导师,研究方向:软件工程。陈星,男,硕士,讲师,研究方向:软件工程。

高,分类及检索方法显著改善,软件开发人员在选择构件时也有了更多的余地,目前大部分评估分类模型的重点为:

- 1) 构件检索中,大部分只关注提高检索的效率——查全率跟查准率。
- 2) 构件选择中,大部分关注点在选择的标准计算上,如可信构件的选择^[2],然后根据相应标准进行构件选择^[3]。

然而在大量的实际工作应用中发现,经过检索后大于阈值的构件往往不止一个,此时有太多选择往往会导致效率低下,并且我们不能仅依赖匹配度去决定一个构件的使用与否,检索到的构件往往是局部最优构件,此构件只对所需功能提供较为完美的实现,但对于整个系统却不尽然,由于检测出的构件功能具有相似性,选出一个最优构件的理由可能会变得模棱两可。此外不同的软件开发人员由于专业知识背景的偏差,以及对所开发系统的了解程度不同,在选择构件时极有可能造成各执己见的场景。全局最优构件直接关系到后期系统的开发的进度和质量,如果选择的构件与系统要求的切合度不高,不仅会导致后期修改工作量巨大,而且可能会推迟开发进度,造成大量人力和财力的损失,使开发陷入困境导致项目失败。考虑到现有的模型几乎只关注构件的评估、分类及检索,对此问题带来的严重危害没有提出有效的解决方法。目前国内外专家仅仅是对该问题可能会导致严重影响有过简单概述,也并未提出一种得到广泛认可的有效处理办法。文献[4]提出了一种针对多构件进行评估和选择的模型,但未给出一种详细的选择方法;文献[5]提出了基于非线性模糊的质量功能展开模型,建立一种基于 NF-QFD 的可复用选择构件模型,但此模型仅把满足用户的需求和系统的可维护性作为考虑的重点,对于构件的全局最优与科学选择阈值问题则未加考虑;文献[6]提出的用法示例代码的构件选择方法虽能帮助用户去了解构件的相关用法,但用户对构件的最终选择仍具有相当大的主观性。

基于上述问题,本文提出了一种结合聚类分析与信息熵进行赋权^[7]的构件选择方法:选取多位专家对构件库系统检索出的构件进行共同判断决策,多位专家结合自己的专业背景知识,在对用户需求进行深入研究了解以后,对每个构件做出评价并得出判断矩阵,凭借每个专家的判断矩阵获得排序向量,使用相关系数法来构造构件相关矩阵,再对聚类阈值劣化数据进行系统分析,从而选出最优的聚类

阈值,对相似程度较高的专家给出合理的聚类,并赋予类间权重(类内权重、类间权重的概念在后文有相关详细定义),数目相对较多的专家所聚得的类分配较大的权重;然后结合信息熵理论分别为类内的专家分配权重,使得同类中认识思维清晰、评价相对准确的分配比较大的权重,最后结合类间权重和类内权重来决定专家的总权重,并通过总权重与构件排序向量计算得出最优构件,此构件才是真正既符合局部需求,又满足整体要求的构件。

2 结合聚类分析与信息熵进行赋权的构件选择方法

2.1 聚类分析理论的概念

所谓聚类分析^[8]是一组将研究对象经过统计计算分为相对同质群组的一种统计分析技术,在聚类的过程中,我们不必给出一个确定分类的标准,聚类分析能够从样本数据出发,通过相关信息自动进行分类,分得的同一个类对象有很大的相似性,不同类对象有较大的差异性。

2.2 结合相关系数的构件聚类分析

2.2.1 1-9 标度法构造构件判断矩阵

假设在某一准则下第 i 个专家对 n 个功能相似的构件进行评价,根据 1~9 标度法^[9]做出构件判断矩阵:

$$A_i = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ \frac{1}{a_{12}} & a_{22} & a_{23} & \cdots & a_{2n} \\ \frac{1}{a_{13}} & \frac{1}{a_{23}} & \vdots & \vdots & a_{3n} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \frac{1}{a_{1n}} & \cdots & \cdots & \cdots & a_{nn} \end{bmatrix}$$

2.2.2 层次分析法构造构件排序向量矩阵

通过层次分析法^[10]中几何平均法的使用对构件的判断矩阵 A_i 进行一致性判断、归一化处理后求出其排序向量 $u_i = (u_1 \ u_2 \ u_3 \ \cdots \ u_n)^T$,其中:

$$u_i = \frac{\omega_i}{\sum_{i=1}^n \omega_i}, \quad i = 1, 2, 3, \cdots, n \tag{1}$$

$$\omega_i = \left(\prod_{j=1}^n a_{ij} \right)^{\frac{1}{n}} \tag{2}$$

则由 m 个专家组成的构件排序向量矩阵为 U :

$$U = (u_1 \ u_2 \ u_3 \ \cdots \ u_m) = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1m} \\ u_{21} & u_{11} & \cdots & u_{2m} \\ u_{31} & u_{32} & \cdots & u_{3m} \\ \vdots & \vdots & \cdots & \vdots \\ u_{n1} & u_{n2} & \cdots & u_{nm} \end{bmatrix}$$

2.2.3 相关系数法构造相关矩阵

由于以上所做分析都是基于相同的量纲,因此所得构件矩阵为标准化矩阵。

接下来确定 $u_i=(u_{1i} \ u_{2i} \ u_{3i} \ \cdots \ u_{ni})^T$ 与 $u_j=(u_{1j} \ u_{2j} \ u_{3j} \ \cdots \ u_{nj})^T$ 之间的相似程度 r_{ij} ,

$$r_{ij}=R(u_i, u_j) \in [0, 1], i, j=1, 2, 3, \cdots, n \quad (3)$$

根据皮尔森相关系数法求得 r_{ij} :

$$r_{ij} = \frac{\sum_{k=1}^n |u_{ki} - \bar{u}_i| |u_{kj} - \bar{u}_j|}{\sqrt{\sum_{k=1}^n (u_{ki} - \bar{u}_i)^2} \cdot \sqrt{\sum_{k=1}^n (u_{kj} - \bar{u}_j)^2}} \quad (4)$$

其中: $\bar{u}_i = \frac{1}{n} \sum_{k=1}^n u_{ki}$, $\bar{u}_j = \frac{1}{n} \sum_{k=1}^n u_{kj}$

由此便可构造一个构件和构件间比较的相关矩阵 R :

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1m} \\ & r_{22} & \cdots & r_{2m} \\ & & \ddots & \vdots \\ & & & r_{mm} \end{bmatrix} \quad (5)$$

2.2.4 基于变化率确定阈值的聚类分析

本文采用聚类分析中的直接聚类法,把 R 中除去对角线以外的元素数据从大到小进行排列,去除重复值得到:

$$1 = \theta_1 > \theta_2 > \theta_3 > \theta_4 > \cdots \quad (6)$$

假设聚类阈值为 $\theta = \theta_k \in [0, 1]$, 当 $r_{ij} > \theta_k (i \neq k)$ 时,我们便认定文 u_i 与 u_j 有高度相似特性,然后将它们划分成同类。由此可以看出,聚类阈值 θ 的取值不同,分类结果可能会大相径庭,而且 θ 越接近 1,分类就越细。如果 $S1$ 与 $S2$ 为通过某一阈值 θ 得出的两个类,若:

$$S1 \cap S2 \neq \emptyset$$

便认定它们是具有相似性,将全部相似的类归并成同类,最终获得的分类便是聚类阈值为 θ_k 时的一种聚类结果,接下来通过变化率选取聚类阈值。

在现实的专家对构件进评价选取中,结果间可能有微小差别,也就是说从专家处获得的构件排序向量的类似程度也许较高,此时我们把阈值 θ 的变化率 C_i 来作为参考进而确定最优阈值:

$$C_i = \frac{\theta_{i-1} - \theta_i}{n_i - n_{i-1}} \quad (7)$$

公式中: i 为 θ 由大到小聚类的次数, n_i 和 n_{i-1} 分别是第 i 次与第 $i-1$ 次聚类的类数量, θ_i 和 θ_{i-1} 是第 i 次与第 $i-1$ 次的聚类阈值,如果:

是第 i 次与第 $i-1$ 次的聚类阈值,如果:

$$C_i = \max(C_1, C_2, C_3, \cdots, C_N) \quad (8)$$

我们就认定第 i 次的聚类阈值为最优阈值。

2.3 结合信息熵的专家赋权分析

在实际的软件选择中,由于每位专家有不同的专业背景知识,不同的履历及偏好,对系统的认识深浅不一,他们的评价结果对评价对象的反映不同,因此我们科学地分配每个专家的权重是十分必要的,

现阶段有两方面的问题需要我们确定:第一是类间权重,类和类之间相比较得出的权重。主要反映每个分得的类里专家的数量,对于有较多专家的类,我们应该给予较大权重;第二是类内权重,即在某个类内,专家间相比较得出的权重。主要根据某个专家分析的逻辑性,清晰度,提供消息的准确性来赋予权值,此时我们需要用信息熵的知识帮助我们更加准确地确定类内的权重。下面分别对类间权重及类内权重提供计算方法。

如果 m 位专家被分成了 t 类 ($t < n$),第 i 类中包含 φ_i ($\varphi_i \leq m$) 位专家, λ_i 为类间权重,则有:

$$\lambda_i = \frac{\varphi_i^2}{\sum_{k=1}^t \varphi_k^2}, i=1, 2, 3 \cdots, t \quad (9)$$

由此我们可以得出每个类的类间权重。

信息熵^[11]一词是 C. E. 香农从热力学里借用的,在热力学中,热熵用来呈现分子状态的混乱水平,香农把信息熵来描述信源的不确定度。对于给定的信源,信息熵应该是一个相应确定的值。在此我们把某位专家对所有可选构件进行评价后的构件排序向量作为一个离散信源,则概率空间为

$$\begin{bmatrix} U \\ P(x) \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \\ P(x_1) & P(x_2) & \cdots & P(x_n) \end{bmatrix} \quad (10)$$

其中: $\sum_{i=1}^n P(x_i) = 1$

m 个专家在既定准则下对 n 个构件进行评价,第 i 位专家通过构件排序向量计算出信息熵为(单位是 BIT):

$$H(U_i) = -\frac{1}{\log_2 n} \cdot \sum_{j=1}^n u_{ij} \log_2(u_{ij}) \quad (11)$$

根据信息熵的表征,我们可以知道,计算出的某位专家的信息熵值越小,信息的不肯定度与随机性就越小,信息确定性越大,表明该专家所做的评估更可信,应该给予更大的权重。因此类内权重可以用信息熵的大小进行度量。

设 I_{ik} 为第 i 类中第 k 位专家类内权重,则:

$$I_{ik} = \frac{[H(U_i)]^{-1}}{\sum_{k=1}^n [H(U_i)]^{-1}} \quad (12)$$

此时 n 代表第 i 类内专家的人数,当专家自成一类时,类内权值定为 1。

到现在,我们已经求得了类间权重与类内权重,接下来对类间权重 λ_i ,类内权重 I_{ik} 相乘,得到第 k 位专家的总权重 β_k ,由此可求出每位专家的总权重。

$$\beta_k = \lambda_i \cdot I_{ik} \quad (13)$$

最后,结合每个专家的总权重和构件排序向量得到最终的构件排序向量 P :

$$P = (P_1 \ P_2 \ \cdots \ P_n)^T$$

$$P_i = \sum_{j=1}^n u_{ij} \cdot \beta_j \quad (14)$$

通过 $\max(P_i)$ 选出最优构件^[12]矩阵。

3 实验与评估

本节给出一个构件选择的实际应用案例,选择对象为从上海构件库中选择一个B/S端,java语言开发的商城系统中文章发布功能模块。该模块包括四个基本功能,查看已发表文章,发表新文章,修改已发表文章,删除已发表文章。经基于剖面检索匹配查找后,大于阈值的构件有四个,由低到高分别为83%、88%、89%、91%。此时四个构件均能较为完善的实现本模块功能,但匹配度为83%的构件与其他三个构件有明显的架构区别,该构件运用了java中的ssh框架和maven进行管理,而其他三个基于spring mvc开发。于此,公司技术部门选出10位

$$U = \begin{bmatrix} 0.5711 & 0.4594 & 0.5882 & 0.5358 & 0.5297 & 0.5495 & 0.6229 & 0.6660 & 0.5384 & 0.5623 \\ 0.2411 & 0.3248 & 0.2244 & 0.3242 & 0.3273 & 0.2905 & 0.2434 & 0.1956 & 0.2515 & 0.2934 \\ 0.1320 & 0.1491 & 0.1334 & 0.0958 & 0.0909 & 0.1013 & 0.0850 & 0.0864 & 0.1655 & 0.1037 \\ 0.0557 & 0.0667 & 0.0540 & 0.0442 & 0.0520 & 0.0587 & 0.0486 & 0.0520 & 0.0446 & 0.0406 \end{bmatrix}$$

因为此矩阵是在相同量纲和数量级下构造,所以此构件排序向量矩阵为标准化矩阵,接了来求相关程度 r_{ij} 。根据式(3)我们可以求得某两位专家

$$R = \begin{bmatrix} 1 & 0.9496 & 0.9989 & 0.9689 & 0.9645 & 0.9875 & 0.9967 & 0.9910 & 0.9946 & 0.9884 \\ & 1 & 0.9343 & 0.9927 & 0.9911 & 0.9792 & 0.9444 & 0.9044 & 0.9598 & 0.9830 \\ & & 1 & 0.9565 & 0.9515 & 0.9796 & 0.9951 & 0.9946 & 0.9922 & 0.9805 \\ & & & 1 & 0.9997 & 0.9949 & 0.9712 & 0.9388 & 0.9671 & 0.9952 \\ & & & & 1 & 0.9937 & 0.9687 & 0.9350 & 0.9609 & 0.9933 \\ & & & & & 1 & 0.9903 & 0.9687 & 0.9808 & 0.9995 \\ & & & & & & 1 & 0.9938 & 0.9835 & 0.9887 \\ & & & & & & & 1 & 0.9739 & 0.9666 \\ & & & & & & & & 1 & 0.9851 \\ & & & & & & & & & 1 \end{bmatrix}$$

工程开发专家对四个构件按匹配度由低到高编号进行评价打分后进行选择,具体步骤如下:

$$A_1 = \begin{bmatrix} 1 & 3 & 5 & 7 \\ \frac{1}{3} & 1 & 2 & 5 \\ \frac{1}{5} & \frac{1}{2} & 1 & 3 \\ \frac{1}{7} & \frac{1}{5} & \frac{1}{3} & 1 \end{bmatrix} \quad A_2 = \begin{bmatrix} 1 & 2 & 3 & 5 \\ \frac{1}{2} & 1 & 3 & 5 \\ \frac{1}{3} & \frac{1}{3} & 1 & 4 \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{4} & 1 \end{bmatrix}$$

$$A_3 = \begin{bmatrix} 1 & 3 & 6 & 7 \\ \frac{1}{3} & 1 & 2 & 4 \\ \frac{1}{6} & \frac{1}{2} & 1 & 5 \\ \frac{1}{7} & \frac{1}{4} & \frac{1}{5} & 1 \end{bmatrix} \quad A_4 = \begin{bmatrix} 1 & 2 & 7 & 8 \\ \frac{1}{2} & 1 & 5 & 6 \\ \frac{1}{7} & \frac{1}{5} & 1 & 4 \\ \frac{1}{8} & \frac{1}{6} & \frac{1}{4} & 1 \end{bmatrix}$$

$$A_5 = \begin{bmatrix} 1 & 3 & 4 & 8 \\ \frac{1}{3} & 1 & 6 & 7 \\ \frac{1}{4} & \frac{1}{6} & 1 & 2 \\ \frac{1}{8} & \frac{1}{7} & \frac{1}{2} & 1 \end{bmatrix} \quad A_6 = \begin{bmatrix} 1 & 2 & 6 & 8 \\ \frac{1}{2} & 1 & 3 & 5 \\ \frac{1}{6} & \frac{1}{3} & 1 & 3 \\ \frac{1}{8} & \frac{1}{5} & \frac{1}{3} & 1 \end{bmatrix}$$

$$A_7 = \begin{bmatrix} 1 & 5 & 6 & 8 \\ \frac{1}{5} & 1 & 4 & 7 \\ \frac{1}{6} & \frac{1}{4} & 1 & 2 \\ \frac{1}{8} & \frac{1}{7} & \frac{1}{2} & 1 \end{bmatrix} \quad A_8 = \begin{bmatrix} 1 & 6 & 7 & 8 \\ \frac{1}{6} & 1 & 3 & 5 \\ \frac{1}{7} & \frac{1}{3} & 1 & 2 \\ \frac{1}{8} & \frac{1}{5} & \frac{1}{2} & 1 \end{bmatrix}$$

$$A_9 = \begin{bmatrix} 1 & 3 & 4 & 7 \\ \frac{1}{3} & 1 & 2 & 6 \\ \frac{1}{4} & \frac{1}{2} & 1 & 6 \\ \frac{1}{7} & \frac{1}{6} & \frac{1}{6} & 1 \end{bmatrix} \quad A_{10} = \begin{bmatrix} 1 & 3 & 6 & 8 \\ \frac{1}{3} & 1 & 4 & 8 \\ \frac{1}{6} & \frac{1}{4} & 1 & 4 \\ \frac{1}{8} & \frac{1}{8} & \frac{1}{4} & 1 \end{bmatrix}$$

由式(1)~(2)计算每个专家的构件排序向量, 则 m 个专家组成的构件排序向量矩阵为 U :

构造的构件排序向量的相关程度,从而构造出形如(5)的相关矩阵 R :

根据式(6)遍历 θ 值并进行聚类,聚类结果如图1~2所示,图1为凝聚计划图,图2为聚类谱系图。

阶段	组合的集群		系数
	集群 1	集群 2	
1	4	5	0.999680
2	6	10	0.999472
3	1	3	0.998926
4	1	7	0.996683
5	4	6	0.995205
6	1	8	0.994598
7	1	9	0.994583
8	2	4	0.992689
9	1	2	0.990296

图1 凝聚计划图

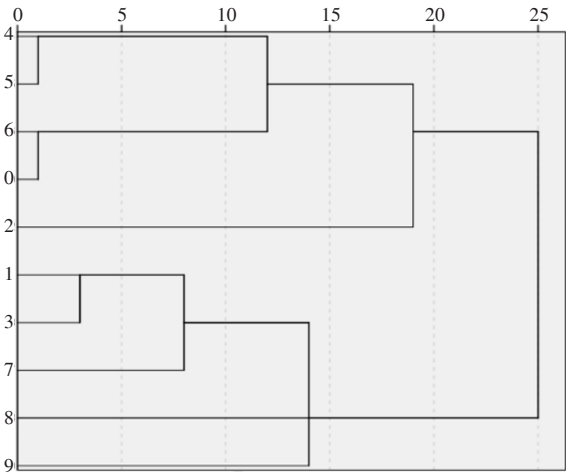


图2 聚类谱系图

根据式(7)可求得每次聚类的变化率:

$$C_0 = \frac{1 - 0.999680}{2 - 0} = 0.000160$$
$$C_1 = \frac{0.999680 - 0.999472}{4 - 2} = 0.000104$$
$$C_2 = \frac{0.999472 - 0.998926}{6 - 4} = 0.000273$$
$$C_3 = \frac{0.998926 - 0.996683}{7 - 6} = 0.002243$$
$$C_4 = \frac{0.996683 - 0.995205}{7 - 7} = \text{无解}$$
$$C_5 = \frac{0.995205 - 0.994598}{8 - 7} = 0.000607$$
$$C_6 = \frac{0.994598 - 0.994583}{9 - 8} = 0.000015$$
$$C_7 = \frac{0.994583 - 0.992689}{10 - 9} = 0.001894$$
$$C_8 = \frac{0.992689 - 0.990296}{10 - 10} = \text{无解}$$

当 C_i 无解、每个专家自成一类或者全部分为一类时没有实际意义,所以剔除 C_4 、 C_8 。根据式(8)可求得 C_3 为最大值,此时 $\theta = 0.996683$,为最优阈值, θ 取此值时,所得聚类为

$$\{(4, 5) (6, 10) (1, 3, 7) (8) (9) (2)\}$$

由此我们便可知分为6类最佳,按上述聚类顺序分别标为

$$t_1, t_2, t_3, t_4, t_5, t_6$$

根据式(9)可算出类间权重,分别为 $t_1=0.2$, $t_2=0.2$, $t_3=0.45$, $t_4=0.05$, $t_5=0.05$, $t_6=0.05$ 。

通过式(11)可计算出各个专家信息熵值,其信息熵值向量表示为

$$(0.78705 \ 0.85615 \ 0.77460 \ 0.77610 \ 0.77465 \ 0.78375 \ 0.71795 \ 0.68895 \ 0.80565 \ 0.75635)$$

根据式(12)可计算出每位专家的类内权重,向量表示为

$$(0.3213 \ 1 \ 0.3265 \ 0.5044 \ 0.4956 \ 0.4911 \ 0.3522 \ 1 \ 1 \ 0.5089)$$

根据式(13)计算出每位专家的总权重,向量表示为

$$(0.1446 \ 0.0500 \ 0.1469 \ 0.1009 \ 0.0991 \ 0.0982 \ 0.1585 \ 0.0550 \ 0.0500 \ 0.1018)$$

最后,结合类位专家的总权重与构件排序向量,根据式(14)得出最终的构件排序向量 P :

$$P = (0.5687 \ 0.2712 \ 0.1115 \ 0.0486)$$

$P_i = \max(P_j)$, 据此我们便能得出,构件1为最优构件,由上述匹配度我们得知,构件1的匹配度相对最低,然而专家中大多数对此构件更加青睐,于是该公司最终选择构件1。后期开发中,基于SSH框架开发的此构件不但简化了开发中业务逻辑处理与数据交互的流程,后期客户又增加的其他功能的编译、依赖问题,maven也能轻松应对,大大节省了开发周期,并避免了不必要的逻辑错误。而如果仅仅按照匹配度选择构件4,则后期必定会出

现二次开发的问题。所以,我们可以说,经过此方法选出的构件,不但结合了专家组们对系统及构件的整体把握,更通过科学有效方法计算出最优,选出的构件能更好地发挥软件复用的优势。

4 结语

本文通过对在构件库中检索出匹配度大于阈值的构件后,从检索结果里再次选择最优构件的研究,提出在实际应用中出现且必须面对的问题,即如何从构件库系统自动检索结果中选出最适合用户需求的构件,并给出结合聚类分析与信息熵进行

Liberation Army.Military software security design guide: GJB/Z 102A-2012 [S]. BeiJing: The General Armament Department military standard publishing department,2012.

[8] 吴继华. 设计与验证 Verilog HDL[M]. 北京:人民邮电出版社出版,2006:77-81.

WU Jihua. Design and verification Verilog HDL[M]. Bei-jing: People post and Telecommunications Press, 2006: 77-81.

[9] 中国人民解放军总装备部电子信息基础部.军用软件质量度量:GJB 5236-2004[S]. 北京:中国标准化出版社,

2004:4-5.

General Armament Department of the Chinese people's Liberation Army. Military software quality metrics: GJB 5236-2004 [S]. Beijing: China Standardization press, 2004:4-5.

[10] 孙梦,宋晓秋,巢翌. 软件程序代码质量度量技术研究 [J]. 计算机工程与设计,2016,27(2):325-327.

SUN Meng, SONG Xiaoqiu, CHAO Yi. Quality measure-ment technology research in software [J]. Computer Engi-neering and Design, 2016, 27(2): 325-327.

(上接第 2441 页)

赋权的一种构件选择方法,此方法不但结合了每位专家的能力水平、经验、客观准确性以及对系统的整体把握等因素,而且结合聚类分析和信息熵理论,达到选出最佳构件的目的。接下来的研究重点在:如何挑选更专业的专家进行构件评价及选择;如何引入其他知识和算法更加科学合理地选择出最佳构件;由于构件为外部人员开发,使用构件便会在一定的风险,如何在评估选择构件时加入风险评估机制也是一个需要考虑的问题。

参 考 文 献

[1] 汪海涛,刘帅,姜瑛,等. 一种基于模糊物元评价法的构件质量度量模型研究[J]. 云南大学学报(自然科学版),2015,37(1):31-42.

WANG Haitao, LIU Shuai, JIANG Ying, et al. Study on a component quality metrics model based on the fuzzy matter-elemenent evaluation method[J]. Journal of Yunnan University, 2015, 37(1): 31-42.

[2] Ding X L, Wang H M, Wang Y Y. Verification oriented trustworthiness evidence and trustworthiness evaluation of software[J]. Journal of Frontiers of Computer Science and Technology, 2010, 4(1):46-48.

[3] Lang B, Liu X D, Wang H M, et al. A classification model for software trustworthiness [J]. Journal of Frontiers of Computer Science and Technology, 2010, 4(3): 231-239.

[4] 盛津芳,王斌. 基于FCD扩展的多构件选择过程[J]. 计算机应用,2007,27(4):860-863.

SHENG Jinfang, WANG Bin. Multi-component Selection Process Based on FCD Extension[J]. Journal of Computer Applications, 2007, 27(4): 860-863.

[5] 孙剑,徐莹. 基于NF-QFD的构件选择方法[J]. 高师理科学刊,2013,33(5):24-28.

SUN Jian, XU Ying. Component selection method based on NF-QFD [J]. Science Journal of Higher Education,

2013, 33(5): 24-28.

[6] 杨秀荣,任姚鹏. 一种基于用法示例的构件选择方法 [J]. 运城学院学报,2013(5):83-85.

YANG Xiurong, REN Yaopeng. A component-based selection method based on usage[J]. Journal of Yuncheng University, 2013(5): 83-85.

[7] 傅祖芸. 信息论[M]. 第三版,北京:电子工业出版社, 2011:132-212.

FU Zuyun. Information Theory [M]. Third Edition, Bei-jing: Publishing House of Electronics Industry, 2011: 132-212.

[8] Bolloju N. Aggregation of analytic hierarchy process models based on similarities in decision makers' preferences [J]. European Journal of Operational Research, 2011, 128 (3):499-508.

[9] 葛世伦. 用1—9标度法确定功能评价系数[J]. 价值工程, 1989(1):33-34.

GE Shilun. With 1-9 scale method to determine the function evaluation coefficient [J]. Value Engineering, 1989 (1): 33-34.

[10] 向东进. 实用多元统计分析[M]. 武汉:中国地质大学出版社,2005:232-256.

XIANG Dongjin. Practical multivariate statistical analysis [M]. Wuhan: China University of Geosciences Press, 2005:232-256.

[11] 陈云翔,董晓雄,项华春,等. 基于信息熵的群组聚类组合赋权法[J]. 中国管理科学,2015,23(6):142-146.

CHEN Yunxiang, DONG Xiongiong, XIANG Huachun, et al. Composition Entropy Weighting Method Based on Information Entropy [J]. Chinese Journal of Management Science, 2015, 23(6): 142-146.

[12] 杨芙清,梅宏. 构件化软件设计与实现[M]. 北京:清华大学出版社,2008:86-145.

YANG Fuqing, MEI Hong. Component-based software design and implementation [M]. Beijing: Tsinghua University Press, 2008: 86-145.