

基于灰色聚类算法的需求分析

胡文生¹ 杨剑锋¹ 赵 明^{2,3}

(贵州理工学院信息工程学院 贵阳 550003)¹ (贵州大学贵州省可靠性工程研究中心 贵阳 550005)²
(瑞典耶夫勒大学技术系 耶夫勒 SE-80176)³

摘 要 James Martin 等人的研究成果表明软件产品的大部分故障来源于需求阶段。为了提高软件产品的质量,保证软件需求规格说明(Software Requires Specification, SRS)的一致性、正确性、无歧义就显得尤为重要。利用自然语言处理技术,通过对软件需求规格说明中功能性需求语句进行词语划分并做词性标注,将每个功能需求语句转换为由关键词构成的权重向量,结合灰色聚类算法最终实现将功能语义相似的需求语句聚合在一起,从而达到需求分类的目的。分类的结果不仅方便需求分析人员进行需求审查,同时也方便软件开发人员进行软件编程以及维护人员对软件系统的修改和变更。

关键词 需求, 软件需求规格说明, 软件故障, 聚类
中图法分类号 TP311 文献标识码 A

Requirements Analysis Based on Grey Clustering Algorithm

HU Wen-sheng¹ YANG Jian-feng¹ ZHAO Ming^{2,3}

(Faculty of Information Engineering, Guizhou Institute of Technology, Guiyang 550003, China)¹
(Reliability Engineering Research Center of Guizhou Province, Guizhou University, Guiyang 550005, China)²
(Department of Technology, University of Gävle, Gävle SE-80176, Sweden)³

Abstract The research results of James Martin et al. show that most of software faults come from the requirements phase. To improve the quality of software products, that the software requirements specification (SRS) must maintain consistency, correctness and unambiguous is very important. The functional requirement statements in SRS are made word division and speech tagging by natural language processing technology. Each functional requirement statement is converted to a weight vector based on the key words and the functional requirement statements with similarity semantics can be clustered by grey clustering algorithm. The results of clustering can not only facilitate the requirements analysis to review SRS, but also facilitate the developer and software maintenance to carry out activities.

Keywords Requires, Software requires specification (SRS), Software fault, Clustering

在软件生命周期中,需求阶段是一个非常重要的阶段。需求阶段所产生的软件需求规格说明(Software Requirements Specification, SRS)是随后设计、开发、测试、维护阶段的基础,如果基础阶段的产品存在问题,必然会影响后续阶段的产品质量。James Martin^[1,2]在一项研究软件产品故障来源的项目中发现,需求阶段引入的故障占到 50% 以上,设计阶段引入的故障占到 27%,而编码阶段引入的故障最少,只占 7%,如图 1 所示。

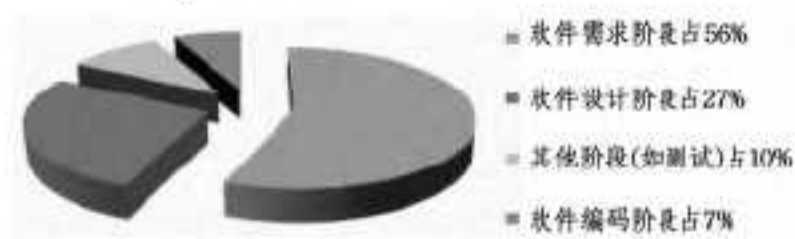


图 1 软件产品故障分布情况

从图 1 中可看出,软件故障主要来源于需求和设计阶段,编程后所引入的各种故障不足 17%。由此可见需求问题一直是软件发生故障的主要来源,也是软件产品需要不断进行

维护的根本原因,这一结论也得到了权威机构 Standish^[3] 的进一步确认。1995 年 Standish 集团在对几百位 IT 执行经理调查研究的基础上形成了最有份量的调研报告《THE STANDISH GROUP REPORT》^[3],这份报告最有价值的地方在于它揭示了软件成功或失败的根本原因,如表 1—表 3 所列。

表 1 软件项目成功的原因

项目成功的因素	所占的百分比(%)
用户积极参与	15.9
执行管理部门的强力支持	13.9
明确、清晰的需求描述	13.0
正确的规划	9.6
切实可行的期望	8.2
项目的重要事件比较少	7.7
称职的员工	7.2
所有权	5.3
有比较清晰的愿景和目标	2.9
勤奋、专注的员工	2.4
其他	13.9

本文受贵州省科技合作计划项目(黔科合 LH 字[2015]7105),贵州省科学技术基金(黔科合 J 字[2015]2064)资助。

胡文生(1969—),男,博士,副教授,主要研究方向为软件可靠性、可维护性,E-mail:1204489713@qq.com;杨剑锋(1987—),男,博士,副教授,主要研究方向为软件可靠性、可维护性,E-mail:332656028@qq.com;赵 明(1962—),男,教授,博士生导师,主要研究方向为软件可靠性、应用统计,E-mail:guidaiiee@hotmail.com。

表 2 项目存在问题的原因

项目存在的问题	所占的百分比(%)
用户投入不足	12.8
不完整的需求及需求规格说明书	12.3
需求及需求规格说明书的变动	11.8
主管部门的支持力度不够	7.5
技术能力不足	7.0
资源不足	6.4
不切实际的期望	5.9
目标不清晰	5.3
不切实际的时限	4.3
新技术的出现	3.7
其他	23.0

表 3 项目失败的原因

项目失败的因素	所占的百分比(%)
不完整的需求	13.1
缺乏用户的参与	12.4
资源不足	10.6
不切实际的期望	9.9
缺乏执行主管部门的支持	9.3
需求及需求规格说明书的经常变动	8.7
缺乏规划	8.1
提供了不再需要的功能	7.5
缺乏软件项目的管理	6.2
对技术一无所知	4.3
其他	9.9%

注:上述数据均来自参考文献[3]。

从表 1—表 3 中可以看出,决定软件项目是否成功的关键因素是需求。在表 1 所列举的成功项目 10 大因素中,有 3 个因素是直接需求有关的,它们是明确且清晰的需求描述及用户积极参与、切实可行的期望,累计所占的百分比为 37.1%;在表 2 对项目存在问题的原因分析中发现,与需求相关的因素分别为用户投入不足、不完整的需求及需求规格说明书、需求及需求规格说明书的变动、不切实际的期望等 4 项,累计所占的百分比为 42.8%;而在表 3 对软件项目失败的 10 大原因分析中直接与需求相关的因素就高达 5 个。由此可见,需求阶段是软件生命周期中最重要的一个阶段,正因为需求很重要,所以有专门的需求工程。在需求工程中有一项非常重要的工作就是需求分析,对于大型的软件项目来说,软件需求成千上万,如果所有的需求都杂乱无章地堆积在一起,不仅会给开发人员、测试人员、维护人员带来困难,而且也会存在相互冲突、相互矛盾的需求,这些存在问题的需求最终会以软件故障的形式表现出来。因此,为了保证软件产品的质量,有必要对软件需求进行正确的分类。本文着重讨论软件需求分析过程中需求分类问题,首先主要介绍一些背景知识,第 1 节阐述了基于灰色聚类算法的软件需求分析,第 2 节进行了实验验证,最后对全文进行总结并对未来的研究方向进行探讨。

1 灰色聚类算法的软件需求分析

软件需求规格说明 (Software Requirements Specification) 主要由两部分组成:非功能性需求和功能性需求。其中非功能性需求主要描述一个软件系统对质量的要求,如系统性能、可维护性、可靠性、安全性等,而功能性需求主要描述软

件系统应该具备的各项功能,在功能性需求之间存在各种各样的关系,如依赖关系、继承关系或语义关系、语法关系等等,本文主要集中在功能性需求语句之间的语义分析上,研究如何通过语义分析来确保软件需求的正确性、一致性、无歧义。通过聚类算法将功能语义相同或相近的需求语句聚合在一起,从而实现对功能性需求语句分类的目的,这种分类有利于指导软件生命周期过程中各个阶段的活动,如需求阶段的验证和确认工作,对于需求分析人员来说,将功能语义相似的需求语句聚合在一起能显著地提高对需求之间是否冲突、是否完整、是否正确的检测效率。对于维护人员来说,利用聚类的结果能有效地实现需求影响分析和变更控制,降低软件需求变更带来的各种负面影响。

对功能性需求语句实施灰色聚类的具体过程如下:

- 1) 利用自然语言处理技术,从软件需求规格说明中每一个功能性需求语句的构词结构入手,从中提炼出代表该需求语句功能语义的关键词,去除重复的关键词以及出现频率比较低的关键词,构成一个关键词集合。
- 2) 根据关键词集合中的每一个关键词在每一个功能性需求语句中的出现情况,构建每个功能性需求语句的权重向量。
- 3) 利用灰色相似关联度和灰色接近关联度的有关理论,构造功能性需求语句之间的相近关联矩阵。
- 4) 借助于统计软件对相近关联矩阵实施聚类分析,从而实现软件需求规格说明中功能性需求语句的分类。

1.1 权重向量的构建

把软件需求规格说明中的功能性需求转化为向量表示,首先从需求规格说明中提炼出关键词,构成一个最终关键词集合。假设从某个软件项目需求规格说明的功能性需求语句中提取出 n 个最终关键词构成的集合为:

$$KeyWord-Set = \{w_1, w_2, \dots, w_n\}$$

则该需求规格说明中的每一个功能性语句可以用如下权重向量来表示:

$$REQ_j = (\frac{w_{j1}}{n_1}, \frac{w_{j2}}{n_2}, \dots, \frac{w_{jn}}{n_n}) \quad (1)$$

其中:

- 1) REQ_j 表示软件需求规格说明中第 j 个功能性需求语句。
- 2) w_{ji} 表示第 i 个关键词在第 j 个功能性需求语句 R_j 中出现的个数。
- 3) n_i 表示第 i 个关键词在全部功能性需求语句中出现的次数。

1.2 计算灰色相近关联矩阵

定义 1^[4,5] 设系统行为序列,

$X_i = (x_i(1), x_i(2), \dots, x_i(n))$ 与 $X_j = (x_j(1), x_j(2), \dots, x_j(n))$ 长度相同, $s_i - s_j$ 如文献[6]所示,则称

$$\epsilon_{ij} = \frac{1}{1 + |s_i - s_j|} \quad (2)$$

为 X_i 与 X_j 的基于相似性视角的灰色关联度,简称相似关联度。

定义 2^[4,5] 设系统行为序列

$X_i = (x_i(1), x_i(2), \dots, x_i(n))$ 与 $X_j = (x_j(1), x_j(2), \dots, x_j(n))$ 长度相同, $S_i - S_j$ 如文献[6]所示,则称

$$\rho_{ij} = \frac{1}{1 + |S_i + S_j|} \quad (3)$$

为 X_i 与 X_j 的基于接近性视角的灰色关联度, 简称接近关联度。

定义 3^[4,5] 设两个序列 X_i 与 X_j 的长度完全相同, ρ_{ij} 和 ε_{ij} 分别是这两个序列的灰色接近关联度和灰色相似关联度, 则称:

$$\gamma_{ij} = \alpha * \varepsilon_{ij} + \beta * \rho_{ij} \quad (4)$$

为序列 X_i 与 X_j 的灰色综合相近关联度, 其中 $\alpha + \beta = 1$ 且 $\alpha, \beta \in [0, 1]$ 。

对所有的 $i \leq j$ 且 $i, j = 1, 2, \dots, m$, 计算出 X_i 与 X_j 的灰色综合相近关联度 γ_{ij} , 得到一个对称矩阵, 取该对称矩阵的上三角矩阵:

$$A = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1m} \\ & \gamma_{22} & \dots & \gamma_{2m} \\ & & \dots & \\ & & & \gamma_{mm} \end{bmatrix} \quad (5)$$

其中, $\gamma_{ii} = 1, i = 1, 2, \dots, m$, 称矩阵 A 为特征变量相近关联矩阵。

1.3 功能需求语句的灰色聚类分析

有两种方法可以实施功能性需求聚类, 一种是取定临界值 $r \in [0, 1]$, 将相近关联矩阵 A 中的每一个数据逐个与临界值 r 比较大小, 对聚类对象进行一次分类, 然后再取一个新临界值 r 与各个子类中的数据比较大小, 从而形成一个更小的分类, 这样一直迭代下去, 形成各级分类方案, 从中选择一个最佳的分类方案即可。这种方法在数据量不是很大的情况下实施起来比较简单, 但是当数据量很大时, 该方法实施起来会有一定的难度。另一种实施聚类的方法是使用统计工具 SPSS 对相近关联矩阵 A 进行自动分析, 从而得到聚类结果。显然在大数据量的情况下第二种方法比第一种方法更好, 本文选用第二种方法。

2 实验验证

以“ATM 自动取款机系统初始需求规格说明书”^[6] 中所包含的 26 个功能性需求语句为例, 依据上述原理实施需求灰色聚类分析。实施步骤过程如下。

1) 构建关键词集合: 首先建立一个空的关键词集合, 记为 $KeyWord-Set = \{\}$ 。由于本文只提取功能性需求语句中表示功能语义的动词以及与动词密切相关的名词, 句中虚词、助词的存在并不影响关键词的提取, 因此建立的停用词表并不包括虚词、助词。对“ATM 自动取款机系统初始需求规格说明书”中 26 个功能性需求语句分析后发现一些并不表示语句功能的动词, 将这些动词作为停用词表: $StopWord-Set = \{\text{“被”、“不能”、“是”、“没有”、“是否”、“都是”、“给”}\}$ 。然后, 对功能性需求语句进行切词分析并进行词性标注, 从中提取表达该语句功能的关键词, “ATM 自动取款机系统初始需求规格说明书”中所包含的关键词集合为: $KeyWord-Set = \{\text{“初始化”、“现金”、“显示”、“取”、“接收”、“消息”、“检查”、“现金卡”、“合法”、“银行代码”、“卡号”、“记录”、“请求”、“口令”、“事务”、“取款”、“金额”、“发送”、“账号”、“安全”}\}$ 。统计关键词集合

中每个关键词出现的次数, 如表 4 所列。

表 4 关键词出现次数统计表

关键词	次数
初始化	2
现金	5
显示	4
取	3
接收	2
消息	7
检查	6
现金卡	10
合法	10
银行代码	4
卡号	3
记录	2
请求	2
口令	8
事务	6
取款	2
金额	3
发送	4
账号	3
安全	1

设置关键词出现频率阈值为 4, 根据上表, 则最终关键词集合为: $Final-Keyword-Set = \{\text{“现金”、“显示”、“消息”、“检查”、“现金卡”、“合法”、“银行代码”、“口令”、“事务”、“发送”}\}$, 共 10 个关键词。

2) 计算每个功能性需求语句的权重向量

最终关键词集合中有 10 个关键词, 因此每个功能性需求语句的权重向量有 10 个分量, 这些分量的取值按照式 (1) 提供的方式进行计算。如计算第一条功能性需求语句的权重向量 REQ_1 的过程如下: 在关键词集合中第一个关键词是“现金”, 而“现金”在表 4 中出现 5 次, 而在该功能性需求语句中出现 0 次, 故该权重向量 REQ_1 第一个分量值为 $0/5 = 0$, 而最终关键词集合 $Final-Keyword-Set$ 中其他关键词在该功能性需求语句中都没有出现, 所以第一条功能性需求语句的权重向量 REQ_1 的其余 9 分量值都为 0。进一步分析, 得出第一条功能性需求语句的权重向量 REQ_1 为:

$$REQ_1 = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$$

其他功能性需求语句的权重向量计算方式与上述一样, 它们分别是:

$$REQ_2 = (1/5, 1/4, 0, 0, 0, 0, 0, 0, 0, 0)$$

$$REQ_3 = (1/5, 1/4, 1/7, 0, 0, 0, 0, 0, 0, 0)$$

$$REQ_4 = (0, 0, 0, 1/6, 2/10, 1/10, 0, 0, 0, 0)$$

$$REQ_5 = (0, 0, 0, 0, 1/10, 1/10, 1/4, 0, 0, 0)$$

$$REQ_6 = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$$

$$REQ_7 = (0, 0, 0, 1/6, 0, 0, 1/4, 2/8, 0, 0)$$

$$REQ_8 = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0)$$

$$REQ_9 = (0, 0, 0, 1/6, 0, 1/10, 0, 1/8, 0, 0)$$

$$REQ_{10} = (0, 1/4, 1/7, 0, 3/10, 0, 0, 1/8, 0, 0)$$

$$REQ_{11} = (0, 0, 0, 0, 0, 0, 0, 1/6, 0, 0)$$

$$REQ_{12} = (0, 0, 0, 0, 0, 0, 0, 1/6, 0, 0)$$

$$REQ_{13} = (0, 0, 0, 0, 0, 0, 0, 1/6, 0, 0)$$

$$REQ_{14} = (1/5, 0, 0, 0, 0, 0, 0, 1/6, 0, 0)$$

$$REQ_{15} = (1/5, 0, 0, 0, 0, 0, 0, 0, 0, 0)$$

$$REQ_{16} = (0, 1/4, 1/7, 0, 1/10, 0, 0, 0, 1/6, 0)$$

$$REQ_{17} = (0, 0, 0, 1/6, 0, 1/10, 1/4, 0, 0, 0)$$

REQ18=(0,0,1/7,0,0,1/10,1/4,0,0,1/4)
 REQ19=(0,0,0,1/6,1/10,1/10,0,1/8,0,0)
 REQ20=(0,0,1/7,0,0,1/10,0,1/8,0,1/4)
 REQ21=(0,0,1/7,0,1/10,2/10,0,1/8,0,1/4)
 REQ22=(0,0,1/7,1/6,1/10,1/10,0,1/8,0,1/4)
 REQ23=(0,0,0,0,0,0,0,0,1/6,0)

REQ24=(1/5,0,0,0,0,0,0,0,0,0)
 REQ25=(0,0,0,0,0,0,0,0,0,0)
 REQ26=(0,0,0,0,0,0,0,0,0,0)

3) 计算灰色关联矩阵

根据式(4)取 $\alpha=\beta=0.5$ 进行编程, 求出上述 26 个权重向量的灰色综合相近关联矩阵(见表 5)。

表 5 $\alpha=\beta=0.5$ 时灰色综合相近关联矩阵

1	0.6041	0.5757	0.6818	0.6897	1	0.6	1	0.7186	0.5501	0.8571	0.8571	0.8571
0.6041	1	0.875	0.5824	0.5891	0.6041	0.5134	0.6041	0.6139	0.4719	0.6526	0.6526	0.6526
0.5757	0.875	1	0.6459	0.6543	0.5757	0.5607	0.5757	0.6847	0.5109	0.6099	0.6099	0.6099
0.6818	0.5824	0.6459	1	0.9836	0.6818	0.8333	0.6818	0.9302	0.7401	0.7692	0.7692	0.7692
0.6897	0.5891	0.6543	0.9836	1	0.6897	0.8219	0.6897	0.9449	0.7311	0.7792	0.7792	0.7792
1	0.6041	0.5757	0.6818	0.6897	1	0.6	1	0.7186	0.5501	0.8571	0.8571	0.8571
0.6	0.5134	0.5607	0.8333	0.8219	0.6	1	0.6	0.7843	0.8687	0.6667	0.6667	0.6667
1	0.6041	0.5757	0.6818	0.6897	1	0.6	1	0.7186	0.5501	0.8571	0.8571	0.8571
0.7186	0.6139	0.6847	0.9302	0.9449	0.7186	0.7843	0.7186	1	0.7012	0.8163	0.8163	0.8163
0.5501	0.4719	0.5109	0.7401	0.7311	0.5501	0.8687	0.5501	0.7012	1	0.6056	0.6056	0.6056
0.8571	0.6526	0.6099	0.7692	0.7792	0.8571	0.6667	0.8571	0.8163	0.6056	1	1	1
0.8571	0.6526	0.6099	0.7692	0.7792	0.8571	0.6667	0.8571	0.8163	0.6056	1	1	1
0.8571	0.6526	0.6099	0.7692	0.7792	0.8571	0.6667	0.8571	0.8163	0.6056	1	1	1
0.6259	0.9231	0.8155	0.5513	0.5572	0.6259	0.4896	0.6259	0.5791	0.452	0.6852	0.6852	0.6852
0.6852	0.8	0.7179	0.4988	0.5036	0.6852	0.4485	0.6852	0.521	0.4172	0.603	0.603	0.603
0.6026	0.5155	0.5633	0.8383	0.8268	0.6026	0.9929	0.6026	0.7887	0.8633	0.6699	0.6699	0.6699
0.6593	0.5633	0.622	0.9524	0.9375	0.6593	0.8696	0.6593	0.8889	0.7685	0.7407	0.7407	0.7407
0.6181	0.5285	0.5791	0.8687	0.8563	0.6181	0.9535	0.6181	0.8155	0.8333	0.6891	0.6891	0.6891
0.6704	0.5727	0.6337	0.9756	0.96	0.6704	0.8511	0.6704	0.9091	0.754	0.7547	0.7547	0.7547
0.6699	0.5722	0.6331	0.9745	0.9589	0.6699	0.8519	0.6699	0.9081	0.7547	0.754	0.754	0.754
0.5907	0.5056	0.5513	0.8155	0.8046	0.5907	0.9745	0.5907	0.7685	0.8889	0.6552	0.6552	0.6552
0.5683	0.487	0.5289	0.7735	0.7636	0.5683	0.915	0.5683	0.7311	0.9449	0.6278	0.6278	0.6278
0.8571	0.6526	0.6099	0.7692	0.7792	0.8571	0.6667	0.8571	0.8163	0.6056	1	1	1
0.6852	0.8	0.7179	0.4988	0.5036	0.6852	0.4485	0.6852	0.521	0.4172	0.603	0.603	0.603
1	0.6041	0.5757	0.6818	0.6897	1	0.6	1	0.7186	0.5501	0.8571	0.8571	0.8571
1	0.6041	0.5757	0.6818	0.6897	1	0.6	1	0.7186	0.5501	0.8571	0.8571	0.8571

(续表)

0.6259	0.6852	0.6026	0.6593	0.6181	0.6704	0.6699	0.5907	0.5683	0.8571	0.6852	1	1
0.9231	0.8	0.5155	0.5633	0.5285	0.5727	0.5722	0.5056	0.487	0.6526	0.8	0.6041	0.6041
0.8155	0.7179	0.5633	0.622	0.5791	0.6337	0.6331	0.5513	0.5289	0.6099	0.7179	0.5757	0.5757
0.5513	0.4988	0.8383	0.9524	0.8687	0.9756	0.9745	0.8155	0.7735	0.7692	0.4988	0.6818	0.6818
0.5572	0.5036	0.8268	0.9375	0.8563	0.96	0.9589	0.8046	0.7636	0.7792	0.5036	0.6897	0.6897
0.6259	0.6852	0.6026	0.6593	0.6181	0.6704	0.6699	0.5907	0.5683	0.8571	0.6852	1	1
0.4896	0.4485	0.9929	0.8696	0.9535	0.8511	0.8519	0.9745	0.915	0.6667	0.4485	0.6	0.6
0.6259	0.6852	0.6026	0.6593	0.6181	0.6704	0.6699	0.5907	0.5683	0.8571	0.6852	1	1
0.5791	0.521	0.7887	0.8889	0.8155	0.9091	0.9081	0.7685	0.7311	0.8163	0.521	0.7186	0.7186
0.452	0.4172	0.8633	0.7685	0.8333	0.754	0.7547	0.8889	0.9449	0.6056	0.4172	0.5501	0.5501
0.6852	0.603	0.6699	0.7407	0.6891	0.7547	0.754	0.6552	0.6278	1	0.603	0.8571	0.8571
0.6852	0.603	0.6699	0.7407	0.6891	0.7547	0.754	0.6552	0.6278	1	0.603	0.8571	0.8571
0.6852	0.603	0.6699	0.7407	0.6891	0.7547	0.754	0.6552	0.6278	1	0.603	0.8571	0.8571
1	0.8571	0.4915	0.5343	0.5032	0.5426	0.5422	0.4826	0.4657	0.6852	0.8571	0.6259	0.6259
0.8571	1	0.4501	0.4851	0.4598	0.4919	0.4915	0.4427	0.4287	0.603	1	0.6852	0.6852
0.4915	0.4501	1	0.875	0.96	0.8563	0.8571	0.9677	0.9091	0.6699	0.4501	0.6026	0.6026
0.5343	0.4851	0.875	1	0.9081	0.9756	0.9767	0.8502	0.8046	0.7407	0.4851	0.6593	0.6593
0.5032	0.4598	0.96	0.9081	1	0.8879	0.8889	0.9302	0.8759	0.6891	0.4598	0.6181	0.6181
0.5426	0.4919	0.8563	0.9756	0.8879	1	0.9988	0.8325	0.7887	0.7547	0.4919	0.6704	0.6704
0.5422	0.4915	0.8571	0.9767	0.8889	0.9988	1	0.8333	0.7895	0.754	0.4915	0.6699	0.6699
0.4826	0.4427	0.9677	0.8502	0.9302	0.8325	0.8333	1	0.9375	0.6552	0.4427	0.5907	0.5907
0.4657	0.4287	0.9091	0.8046	0.8759	0.7887	0.7895	0.9375	1	0.6278	0.4287	0.5683	0.5683
0.6852	0.603	0.6699	0.7407	0.6891	0.7547	0.754	0.6552	0.6278	1	0.603	0.8571	0.8571
0.8571	1	0.4501	0.4851	0.4598	0.4919	0.4915	0.4427	0.4287	0.603	1	0.6852	0.6852
0.6259	0.6852	0.6026	0.6593	0.6181	0.6704	0.6699	0.5907	0.5683	0.8571	0.6852	1	1
0.6259	0.6852	0.6026	0.6593	0.6181	0.6704	0.6699	0.5907	0.5683	0.8571	0.6852	1	1

4) 功能性需求聚类分析

目前国内外使用最多的聚类方法是系统聚类方法, 该方法的基本思想是^[7,8], 先将 n 个样品各自看成一类, 然后规定

样品之间的距离和类与类之间的距离, 刚开始由于每个样品自成一类, 类与类之间的距离与样品之间的距离是相等的, 选择距离最小的一对合并成一个新类, 计算新类和其他类的距

离,再将距离最近的两类合并,这样每次减少一类,直至所有的样品都成一类为止。根据类与类之间的距离定义方式的不同产生了不同的系统聚类方法,其中广泛使用的方法是类平均法和离差平方和法(有时也称 Ward 法)。类平均法就是将两类之间的距离平方定义为这两类元素两两之间的平均平方距离,而离差平方和法的思想来自于方差分析,如果类分得正确,同类样品的离差平方和应当较小,类与类之间的离差平方和应当较大。研究需求聚类的目的是使功能语义相似的需求语句聚合在一起,所以选用 Ward 法,使用 SPSS 工具对表 5 中的数据执行聚类分析,得到的结果如图 2 所示。

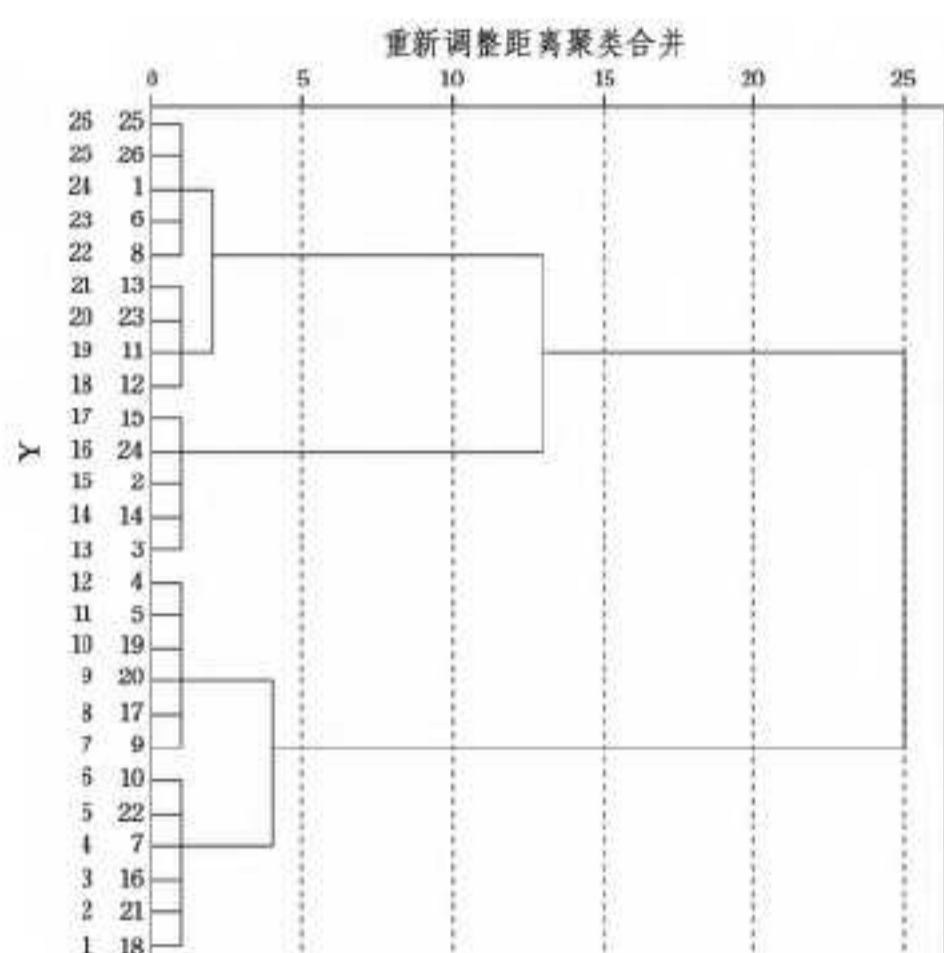


图 2 聚类分析结构树状图

得到聚类树状图后,还需要确定究竟分成多少类比较合理。从图 2 来看,可供分类的方案有 5 类、4 类、3 类、2 类。一般情况下,在聚类过程中希望每个类别包含的对象数基本相等,因此找出尽可能“等分”的分类就是通常确定类别数目的原则之一^[8,9]。结合图 2,可以统计出 26 个功能性需求语句在分为 2~5 类时各个类别的需求数,如表 6 所列。从表中可以发现,分为 5 类相对比较合适,这时最大类别数为 6,占全部功能性需求语句数的 23.1%,最小的类别数为 4,占全部功能性需求语句数的 15.4%,分类情况相对均匀。

表 6 不同分类情况下各个类别包含的需求数

类别	1	2	3	4	5
Ward Method(5)	5	4	5	6	6
Ward Method(4)	9	5	6	6	
Ward Method(3)	9	5	12		
Ward Method(2)	14	12			

26 个功能性需求的分类情况如表 7 所列。

表 7 功能性需求语句分类情况

类别	类别中的成员	类别特征描述
1	{R25, R26, R1, R6, R8}	该组由对 ATM 初始化、安全保证方面的需求语句组成。
2	{R13, R23, R11, R12}	该组主要由 ATM 处理事务方面的需求语句组成。
3	{R15, R24, R2, R14, R3}	该组主要由与提取现金相关的需求语句构成。
4	{R4, R5, R19, R20, R17, R9}	该组由关于合法性检查方面的需求语句组成。
5	{R10, R22, R7, R16, R21, R18}	该组主要由描述 ATM 与银行计算机交互情况的需求语句构成。

结束语 需求阶段是软件生命周期过程中最为重要的阶段,该阶段所产生的产品——软件需求规格说明是其它各个阶段开展工作的基础,如果基础坏了,那么软件项目就会失败,因此需求分析阶段必须严格管理。对于一个大型的软件项目来说,需求可能会成千上万,如果杂乱无章地堆积在一起,会给开发过程和今后的维护带来困难。本文通过自然处理技术及灰色聚类算法,将需求按照某些标准和原则进行分类,从而实现将功能相近的需求放在一起,这无论是对需求审查,还是今后的软件开发、软件维护,都会有很大的帮助。

参考文献

- [1] Martin J. An Information Systems Manifesto[M]. Upper Saddle River, NJ, USA: Prentice Hall PTR, 1986
- [2] Hu Wen-sheng, Zhao Ming, Wu Shi-yi, et al. Requires analysis based on software maintainability[C]// 2014 International Conference Reliability, Maintainability and Safety(ICRMS). 2014: 354-357
- [3] 徐锋. 软件需求最佳实践——SERU 过程框架原理与应用[M]. 北京: 电子工业出版社, 2011: 2-15
- [4] 郭晓君, 刘思锋, 方志耕. 基于合成灰数灰度的区间灰数自忆性预测模型[J]. 系统工程与电子技术, 2014, 36(6): 1124-1129
- [5] 刘思峰, 谢乃明, 等. 灰色系统理论及其应用[M]. 北京: 科学出版社, 2008: 52-63
- [6] 余金山. 实时 UML Rational Rose RealTime 建模案例剖析[M]. 北京: 电子工业出版社, 2007
- [7] 张尧庭, 方开泰. 多元统计分析引论[M]. 武汉: 武汉大学, 2013, 32(6): 1071-1078
- [8] 戴晖, 郭均鹏, 高飒, 等. 一般分布区间型符号数据的系统聚类[J]. 数理统计与管理, 2013, 32(6): 1071-1078

(上接第 451 页)

- [16] Yamaguchi O, Fukui K, Maeda K. Face recognition using temporal image sequence[C]// Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, Nara, 1998: 318-323
- [17] Fukui K, Yamaguchi O. Face recognition using multi-viewpoint patterns for robot vision[C]// Proceedings of the International Symposium of Robotics Research, Siena, Italy, 2003: 192-201
- [18] Nishiyama M, Yamaguchi O, Fukui K. Face Recognition with the multiple constrained mutual subspace method[C]// Proceedings of the 5th International Conference on Audio and Video-Based Biometric Person Authentication, New York, 2005: 71-80
- [19] Li J W, Wang Y H, Tan T N. Video-based face recognition using

- a metric of average Euclidean distance[C]// Proceedings of the 5th Chinese Conference on Biometric Recognition, Guangzhou, China, 2004: 224-232
- [20] Li J W, Wang Y H, Tan T N. Video-based face recognition using earth mover's distance[C]// Proceedings of the International Conference on Audio and Video-based person Authentication, New York, 2005: 229-239
- [21] Fan W, Wang Y H, Tan T N. Video-based face recognition using Bayesian inference model[C]// Proceedings of the International Conference on Audio and Video-based Person Authentication, New York, 2005: 122-130
- [22] 宛根训, 田强, 李建勇, 等. 公安大数据背景下的人脸识别技术进展[J]. 警察技术, 2014(5): 16-18