



# 中华人民共和国国家标准

GB/T 29836.2—2013

## 系统与软件易用性 第2部分：度量方法

Usability of software system—Part 2: Metric method

2013-11-12 发布

2014-02-01 实施

中华人民共和国国家质量监督检验检疫总局  
中国国家标准化管理委员会 发布

目 次

前言 ..... I

引言 ..... II

1 范围 ..... 1

2 规范性引用文件 ..... 1

3 术语和定义 ..... 1

4 概述 ..... 1

5 易理解性 ..... 1

6 易学习性 ..... 2

7 易操作性 ..... 3

8 吸引力 ..... 5

9 易用性评价方法 ..... 6

    9.1 概述 ..... 6

    9.2 确定评价因素、评价等级..... 6

    9.3 构造评判矩阵和确定权重 ..... 7

    9.4 进行模糊合成和作出决策 ..... 7

附录 A（资料性附录） 网上评卷系统易用性评价示例 ..... 9

参考文献 ..... 27

## 前 言

GB/T 29836 在《系统与软件易用性》总标题下,目前分为如下三部分:

——第1部分:指标体系;

——第2部分:度量方法;

——第3部分:测评方法。

本部分为 GB/T 29836 的第2部分。

本部分按照 GB/T 1.1—2009 给出的规则起草。

本部分由全国信息技术标准化技术委员会(SAC/TC 28)提出并归口。

本部分起草单位:中国电子技术标准化研究院、北京邮电大学、南宁市平方软件新技术有限责任公司、国家应用软件产品质量监督检验中心、教育部考试中心、上海计算机软件技术开发中心、上海浦东软件平台有限公司、上海宝信软件股份有限公司、珠海南方软件网络评测中心、上海鲁齐信息科技有限公司。

本部分主要起草人:张旻旻、袁玉宇、李英华、朱洁、郭新伟、胡宇、蔡立志、吴玉军、杨金翠、温家凯、韩强、侯建华、左家平、丁志刚、李家宏、张露莹、崔岩。

GB/T 29836.2—2013

## 引 言

GB/T 29836 的本部分提出了易用性质量度量方法。本部分将人(参与测试的用户)、机(计算机系统)及人机交互作为计算机系统与软件的三大要素,综合三大要素对被测目标系统易用性的影响。

本部分对已经构建设计的系统与软件易用性指标体系中的每个测试指标,给出详细的度量目的、度量描述、度量公式以及度量公式中各个度量元素的描述说明。并对度量的最终结果值,给出解释。本部分适用于各种具有人机交互的计算机软件产品及相关系统。

本标准预期的主要使用者包括:

- a) 需方(从供方获得或采购系统、软件产品或软件服务的个体或组织);
- b) 独立评价者(与软件无利益关系、受委托实施独立评价的个体或组织);
- c) 开发者(执行开发活动的个体或组织);
- d) 维护者(执行维护活动的个体或组织);
- e) 供方(按所签合同向需方提供系统、软件产品或软件服务的个体或组织),其在合格性测试中确认软件质量时使用;
- f) 用户(使用软件产品执行具体功能的个体或组织),其在验收测试中评价软件产品质量时使用;
- g) 质量管理者(执行软件产品或软件服务的系统性检查的个体或组织),作为质量保证和质量控制的组成部分评价软件质量时使用。

GB/T 29836.1—2013《系统与软件易用性 第1部分:指标体系》给出了易用性指标体系,GB/T 29836.3—2013《系统与软件易用性 第3部分:测评方法》描述了易用性指标的测评方法。本部分旨在与 GB/T 29836.1—2013 和 GB/T 29836.3—2013 联合使用。

## 系统与软件易用性 第2部分：度量方法

### 1 范围

GB/T 29836 的本部分在 GB/T 29836.1 提出的指标体系的基础上,规定了系统与软件的易用性度量公式,为系统与软件的供方、需方、第三方评测机构提供统一的易用性度量方法。

本部分适用于具有人机交互的各类系统与软件,本部分不适用于无人交互功能或者在使用期间无易用性需求的系统与软件。

注：本部分中所指的系统主要是软件系统。

### 2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB/T 11457 软件工程术语

GB/T 16260.1 软件工程 产品质量 第1部分：质量模型 (GB/T 16260.1—2006,ISO/IEC TR 9126-1:2001,IDT)

GB/T 29836.1—2013 系统与软件易用性 第1部分：指标体系

### 3 术语和定义

GB/T 29836.1—2013、GB/T 16260.1 和 GB/T 11457 中界定的术语和定义适用本文件。

### 4 概述

易用性是系统与软件的一个重要质量特征。对系统与软件的易用性进行度量,有利于了解软件是否满足规定的易用性要求,有利于需方、供方等及时发现、定位易用性的缺陷。

在实施度量之前,宜预先申明相关度量对象(如任务、功能)的粒度。

本部分提出了与 GB/T 29836.1 一起使用的一组软件易用性质量的度量。用户宜从 GB/T 29836.1 中选择易用性指标,根据指标确定度量以及相应的测试方法。对于某一个具体的系统或软件,不同模块或功能,易用性有不同的要求。因此可以根据其易用性的具体目标选取相应合适的指标,提高易用性度量的准确性。对于与易用性完全不相关的功能点,度量时可以放弃。对于其他无法度量的指标,度量时可以放弃。

### 5 易理解性

易理解性具体指标的度量方法如表 1 所示。

GB/T 29836.2—2013

表 1 易理解性

指标名称	指标描述	测量、公式及数据元素计算	测量值说明
明显的功能	基于初始的条件,产品的功能(或功能的类型)能被用户识别的比例是多少	$X=A/B$ 式中: A——在培训前,用户可以根据界面提示进行操作的功能数; B——在培训前,用户应该根据界面提示进行操作的功能数	$0\leq X\leq 1$ ,越接近 1,表明该项指标越好
描述的完整性	在阅读完用户手册后能正确理解的功能(或功能的类型)比例是多少	$X=A/B$ 式中: A——在阅读完用户手册后能正确理解的功能(或功能的类型)数; B——用户需要理解的功能(或功能的类型)数	$0\leq X\leq 1$ ,越接近 1,表明该项指标越好
演示能力	需要演示的功能中具有演示能力的比例是多少	$X=A/B$ 式中: A——在需要演示的功能中具有演示能力的功能数; B——需要演示的功能的数	$0\leq X\leq 1$ ,越接近 1,表明该项指标越好
演示的有效性	在演示或指导之后用户能成功执行功能的比例是多少	$X=A/B$ 式中: A——在演示或指导之后用户能成功执行功能的数; B——用户本应需要执行的功能数	$0\leq X\leq 1$ ,越接近 1,表明该项指标越好
输入的有效性检查	输入项提供了对有效数据进行检查的比例是多少	$X=A/B$ 式中: A——能够对输入数据进行有效性检查的功能数; B——要求对输入数据进行有效性检查的功能数	$0\leq X\leq 1$ ,越接近 1,表明该项指标越好

6 易学习性

易学习性具体指标的度量方法如表 2 所示。

表 2 易学习性

指标名称	指标描述	测量、公式及数据元素计算	测量值说明
帮助文档的有效性	在阅读完帮助文档后能理解的功能(或功能的类型)比例是多少	$X=A/B$ 式中： A——用户在使用阅读完帮助文档后理解的功能数； B——软件的功能数	$0\leq X\leq 1$ ，越接近 1，表明该项指标越好
帮助机制的有效性	在使用了帮助后，能正确地完成任务的比例是多少	$X=A/B$ 式中： A——用户使用帮助后完成的功能数； B——用户需要使用帮助完成的功能数	$0\leq X\leq 1$ ，越接近 1，表明该项指标越好

7 易操作性

易操作性具体指标的度量方法如表 3 所示。

表 3 易操作性

指标名称	指标描述	测量、公式及数据元素计算	测量值说明
使用中默认值的可用性	为便于操作，用户能否易于选择参数值	$X=A/B$ 式中： A——用户可自己选择的参数的功能数； B——要求可供用户自己选择参数的功能数	$0\leq X\leq 1$ ，越接近 1，表明该项指标越好
完成指定任务的步骤	用户完成指定任务所用的步骤数目	$X=A/B$ 式中： A——符合完成所用的步骤数要求的任务数； B——要求必须进行评测的任务数目	$0\leq X\leq 1$ ，越接近 1，表明该项指标越好
操作的复杂性	用户完成指定任务总体复杂度	$X=A/B$ 式中： A——用户完成指定任务的动作数； B——要求用户完成指定任务所要求的动作数	$0\leq X\leq 1$ ，越接近 1，表明该项指标越好
完成指定任务过程中误操作的次数	用户完成指定任务所产生的误操作数	$X=A/B$ 式中： A——用户在使用系统完成某项任务时出现误操作的步骤次数； B——用户在正常情况下，完成某项任务所用的步骤次数	$0\leq X\leq 1$ ，越接近 1，表明该项指标越好



GB/T 29836.2—2013

表 3 (续)

指标名称	指标描述	测量、公式及数据元素计算	测量值说明
错误的纠正	用户能否容易地纠正完成任务过程中出现的错误	$X = A/B$ 式中： A——用户在使用系统完成某项任务时成功纠正错误的次数； B——用户在使用系统完成某项任务时出现错误的次数	$0 \leq X \leq 1$ , 越接近 1, 表明该项指标越好
发生错误的影响力	用户发生错误之后, 对完成任务的影响程度	$X = A/B$ 式中： A——用户在完成任务过程中出现误操作后, 完成任务所需要的时间； B——用户使用系统完成某项任务的完成时间	$1 \leq X \leq \infty$ , 越接近 1, 表明该项指标越好
可还原性	用户正确地还原到操作之前状态的能力如何	$X = A/B$ 式中： A——可以恢复到原状态的功能数； B——要求可以恢复到原状态的功能数	$0 \leq X \leq 1$ , 越接近 1, 表明该项指标越好
运行差错的易恢复性	能够容忍用户差错并帮助用户恢复的功能的比例是多少	$X = A/B$ 式中： A——能够容忍用户差错并帮助用户恢复的功能的数； B——软件的功能数	$0 \leq X \leq 1$ , 越接近 1, 表明该项指标越好
使用中的消息的可理解性	用户是否容易理解软件系统的消息？用户是否容易记住重要的消息 在开始下一步动作之前是否有任何引起用户延缓理解的消息	$X = A/B$ 式中： A——在使用过系统后, 能够被用户理解的系统消息数； B——使用系统过程中提示的消息数	$0 \leq X \leq 1$ , 越接近 1, 表明该项指标越好
运行状态的易监控性	具有运行状态监控能力的功能的比例是多少	$X = A/B$ 式中： A——能够向用户提供运行状态查看的功能数目； B——要求向用户提供运行状态查看的功能数目	$0 \leq X \leq 1$ , 越接近 1, 表明该项指标越好
界面元素的易定制性	是否提供多种风格或方案供用户选择	$X = A/B$ 式中： A——用户可以选择多种风格或方案的功能数； B——要求用户可以选择多种风格或方案的功能数	$0 \leq X \leq 1$ , 越接近 1, 表明该项指标越好



表 3（续）

指标名称	指标描述	测量、公式及数据元素计算	测量值说明
界面布局的易定制性	用户能否为方便自己定制界面的布局	$X=A/B$ 式中： A——用户可以改变的界面布局元素数； B——要求用户可以改变的界面元素数	$0\leq X\leq 1$ ，越接近 1，表明该项指标越好
快捷方式的易定制性	用户是否能方便地定制快捷方式	$X=A/B$ 式中： A——用户可以自定义快捷键的功能数； B——要求用户可以自定义快捷键的功能数	$0\leq X\leq 1$ ，越接近 1，表明该项指标越好
操作规程的易定制性	用户能否为方便自己容易地定制其操作规程	$X=A/B$ 式中： A——用户可以自定义操作规程的功能数； B——要求用户可以自定义操作规程的功能数	$0\leq X\leq 1$ ，越接近 1，表明该项指标越好
特殊辅助功能	提供辅助功能的种类	$X=A/B$ 式中： A——已经提供辅助功能的种类数； B——要求提供辅助功能的种类数	$0\leq X\leq 1$ ，越接近 1，表明该项指标越好
无障碍程度	能让有身体障碍的用户操作的功能比例是多少	$X=A/B$ 式中： A——能让有身体障碍的用户操作的功能的数； B——软件的功能数	$0\leq X\leq 1$ ，越接近 1，表明该项指标越好
一致性	系统的各部分之间以及与系统之外的相关因素是否保持一致	$X=A/B$ 式中： A——符合内、外部一致性要求的系统元素数； B——要求必须符合内部一致性要求的系统元素的数	$0\leq X\leq 1$ ，越接近 1 越好，表明该系统一致性越好

8 吸引性

吸引性具体指标的度量方法如表 4 所示。

表 4 吸引性

指标名称	指标描述	测量、公式及数据元素计算	测量值说明
界面色彩对视觉的吸引性	色彩对用户有多大的吸引力、是否让用户感到视觉舒适？从而减少视觉工作量	$X=A/B$ 式中： A——在使用过系统后，感觉界面配色方案舒适的用户人数； B——使用过系统的用户人数	$0\leq X\leq 1$ ，越接近 1，表明该系统的配色方案舒适度越好，越被用户所接受
界面元素形状的舒适度	界面各种元素的形状是否舒适	$X=A/B$ 式中： A——在使用过系统后，感觉界面形状方案舒适的用户人数； B——使用过系统的用户人数	$0\leq X\leq 1$ ，越接近 1 越好，表明该系统的形状方案舒适度越好，越被用户所接受
界面元素尺寸的合理性	界面各种元素的尺寸是否合理	$X=A/B$ 式中： A——在使用过系统后，感觉界面尺寸方案舒适的用户人数； B——使用过系统的用户人数	$0\leq X\leq 1$ ，越接近 1 越好，表明该系统的尺寸方案舒适度越好，越被用户所接受
布局的合理性	界面的各种元素的配合是否合理 能否减少记忆工作	$X=A/B$ 式中： A——符合工作要求及操作方便性要求的界面布局及元素数； B——要求必须符合工作要求及操作方便性要求的界面布局及元素数	$0\leq X\leq 1$ ，越接近 1 越好，表明该系统界面布局越符合工作要求且操作方便性好
用户的感受度	用户对软件直观感觉的综合评价	$X=A/B$ 式中： A——对使用用户进行调查，认为感受度高的用户； B——进行调查的所有用户	$0\leq X\leq 1$ ，越接近 1 越好，表明该系统用户感受度好

9 易用性评价方法

9.1 概述

易用性评价宜使用模糊综合评价法。该方法根据模糊数学的隶属度理论把定性评价转化为定量评价，即用模糊数学对受到多种因素制约的事物或对象做出一个总体的评价。运用模糊综合评价法进行决策时有三个步骤：1)确定评价因素、评价等级；2)构造评判矩阵和确定权重；3)进行模糊合成，作出决策。其中，评价因素是对指标评议的具体内容，评价等级是评价因素的优劣程度，被评价因素确定了一个从指标到等级的模糊关系矩阵以及权重，权重是表示指标相对重要性大小的量度值。

9.2 确定评价因素、评价等级

设  $U=\{u_1,u_2,\cdots,u_m\}$  为被评价对象的  $m$  种评价指标；

$V=\{v_1,v_2,\cdots,v_n\}$  为每一指标所处的评价等级。

这里  $m$  为评价指标的个数,  $n$  为评价的个数。

这些指标可以衡量被测系统的易用性“水平”。评测结果的评定按四级评分, 设定优秀、良好、及格、不及格。评判等级应尽量满足人们区分易用性水平的要求和一般习惯。

### 9.3 构造评判矩阵和确定权重

首先着眼因素集中的单因素  $u_i (i=1, 2, \dots, m)$  作单指标评判, 从指标  $u_i$  着眼该事务对抉择等级  $v_j (j=1, 2, \dots, n)$  的隶属为  $r_{ij}$ , 得出第  $i$  个指标  $u_i$  的单指标评判集:

$$r_i = (r_{i1}, r_{i2}, \dots, r_{in})$$

$m$  个着眼因素的评价集构造一个总的评价矩阵  $R$ 。即每一个被评价系统确定了从  $U$  到  $V$  的模糊关系  $R$ , 它是一个矩阵:

$$R = (r_{ij})_{m \times n} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ r_{m1} & r_{m2} & \cdots & r_{mn} \end{bmatrix}, (i=1, 2, \dots, m; j=1, 2, \dots, n)$$

其中  $r_{ij}$  表示从指标  $u_i$  着眼, 该评测系统能被评为  $v_j$  的隶属度。即  $r_{ij}$  表示第  $i$  个指标  $u_i$  在第  $j$  个评语  $v_j$  上的频率分布, 一般将其归一化使之满足  $\sum r_{ij} = 1$ , 这样  $R$  阵本身就是没有量纲的, 不需作专门处理。

用等级比重确定隶属矩阵的方法, 可以满足量化综合评价的要求。用等级比重法确定隶属度时, 为了保证可靠性, 要注意两个问题: 第一, 评价者人数不可太少, 因为只有这样, 等级比重才趋于隶属度; 第二, 评价者必需对被评测系统有相当的了解, 特别是对于业务系统易用性的评测, 更应该如此。

得到这样的模糊关系矩阵, 尚不足以对系统易用性作出评测。评价指标集中的各个指标在“评价目标”中有不同的地位和作用, 即对评价系统在综合评价中有不同的比重。拟引入  $U$  上的一个模糊子集  $A$ , 称权重分配集:

$$A = (a_1, a_2, \dots, a_m), \text{ 其中 } a_i \geq 0, \text{ 且 } \sum a_i = 1. \text{ 它反映对诸指标的一种权衡。}$$

在评价系统中, 权重的确定是至关重要的。常见评价问题中的权重值, 一般多凭经验主观给定, 成为主观赋权。主观确定权重也有客观的一面, 一定程度上反映了实际情况, 评价的结果有较高的参考价值。确定权重也可以利用数学的方法, 数学方法严格的逻辑性而且可以对确定的“权重”进行“滤波”和“修复”处理, 以尽量剔除主观成分, 符合客观事实。

这里存在两种模糊集, 以主观赋权为例, 一类是标志指标集  $U$  中各指标在人们心目中的重要程度的量, 表现为指标集  $U$  上的模糊权重向量  $A = (a_1, a_2, \dots, a_m)$ ; 另一类是  $U \times V$  上的模糊关系, 表现为  $m \times n$  模糊矩阵  $R$ 。这两类模糊集都是人们对系统易用性印象的反映。

### 9.4 进行模糊合成和作出决策

$R$  中不同的行反映系统易用性从不同的指标来看对各等级模糊子集的隶属程度。用模糊权向量  $A$  将不同的行进行综合, 即得到被评系统易用性从总体上来看对各等级模糊子集的隶属程度, 即模糊综合评价结果向量。

引入评价集  $V$  上的一个模糊子集  $B$ , 称模糊评价, 又称决策集,  $B = (b_1, b_2, \dots, b_n)$ 。

通过  $R$  与  $A$  求  $B$  的方法是, 令  $B = A * R$  ( $*$  为算子符号), 称为模糊变换。

这个模型看起来很简单, 但实际上较为复杂, 对于不同模糊算子, 就不同的评价模型。

$A$  称输入,  $B$  称输出。

如果评判结果  $\sum b_j \neq 1$ , 应将它归一化。

$b_j$  表示被评价系统易用性具有评语  $v_j$  的程度。各个评判指标, 具体反映了被评系统易用性在所评判的特征方面的分布状态, 使评判者对评判对象有更深入的了解, 并能作各种灵活的处理。如果要选择

一个决策,则可选择最大的  $b_j$  所对应的等级  $v_j$  作为综合判断的结果。

$B$  是对每个评测指标综合状况分等级的程度表述,它不能直接用于被评判系统间的排序评优,更需要更进一步的分析处理,待分析处理之后才能应用。采用最大隶属度法则对其处理,得到最终评判结果。此时,我们只利用了  $b_j (j=1,2,\cdots,n)$  中的最大者,没有充分利用  $B$  所带来的信息。为了充分利用  $B$  所带来的信息,可把各种等级的评级参数和评判结果  $B$  进行综合考虑,使得评判结果更加符合实际。

设相对于各等级  $v_j$  规定的参数列向量为:

$$C=(c_1,c_2,\cdots,c_n)^T$$

则得出等级参数评判结果为:

$$B * C = p$$

$p$  是一个实数。它反映了等级模糊子集  $B$  和等级参数向量所带来的综合信息,在许多实际应用中,它是十分有用的综合参数。

就理论而言,上述的模糊合成运算有无穷多种,但在实际应用中,经常采用具体模型有几种。关于  $B$  的求法,最早的合成运算采用查德算子(主因素突出型)。但当评价因素较多时,由于  $a_i$  很小,评判结果得到的  $b_j$  反映不出实际情况,失掉了综合评价的意义。因此,应用查德算子作综合评判,往往得到的结果与实际情况相差很大。为了克服这一缺点,根据实际情况采用其他类型的“与”、“或”算子,或者将两种类型的算子搭配使用。较简单的是普通矩阵乘法(即加权平均法),这种模型要让每个因素都对综合评价有所贡献,比较客观地反映了评测系统易用性的全貌。只要采用的算子对一方面抓住实际问题的本质,获得满意的效果;另一方面保证满足  $0 \leq b_j < 1$ 。

附录 A 给出了网上评卷系统易用性评价示例。



附 录 A  
(资料性附录)  
网上评卷系统易用性评价示例

A.1 引言

网上评卷系统易用性评测体系作为本标准在实际应用中的案例,该体系分为指标体系、测评方法、评价方法三部分。本附录将在 GB/T 29836.1—2013 和 GB/T 29836.3—2013 的基础上,给出网上评卷系统易用性的评价方法。指标体系的选取和构建过程、测评方法分别在 GB/T 29836.1—2013 的附录 B 和 GB/T 29836.3—2013 的附录 A 中进行详细分析与介绍。

A.2 网上评卷系统易用性评价指标体系

在网上评卷系统易用性指标体系在 GB/T 29836.1—2013 中已经给出,在这里我们需要将该体系中的特性和评测指标赋予到相应的评价指标集合中。

网上评卷易用性评测指标体系是四层结构的,第一层为总目标层,即网上评卷系统易用性( $U$ ),第二层为产品指标( $P$ )和交互指标( $I$ ),第三层为产品指标( $P$ )和交互指标( $I$ )的指标类别,第四层为具体指标。最终的网上评卷易用性评价体系。如图 A.1 所示。

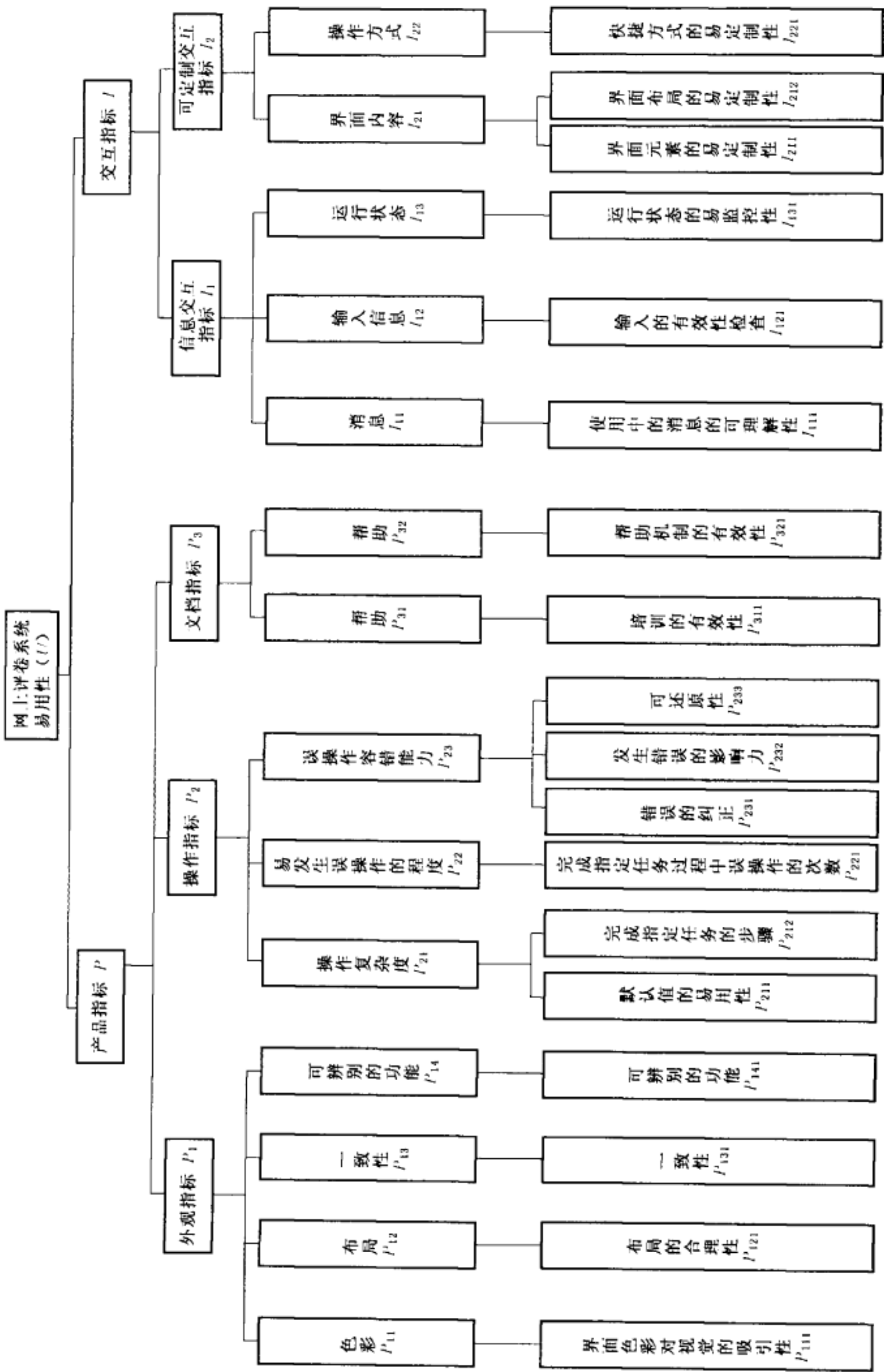


图 A.1 网上评卷易用性评价体系



网上评卷系统易用性评价指标集分成四个层次:

- a) 总目标层集:  $U=(P, I)$ , 即总目标, 网上评卷系统易用性 = {产品指标, 交互指标}。
- b) 第二级因素层: 即第二层指标集:
  - 1)  $P=(P_1, P_2, P_3)$ , 产品指标 = {外观指标, 操作指标, 文档指标};
  - 2)  $I=(I_1, I_2)$ , 交互指标 = {信息交互指标, 可制定交互指标}。
- c) 第三级因素层: 即第三层指标集:
  - 1)  $P_1=(P_{11}, P_{12}, P_{13}, P_{14})$ , 外观指标 = {色彩, 布局, 一致性, 明显的功能};
  - 2)  $P_2=(P_{21}, P_{22}, P_{23})$ , 操作指标 = {操作复杂度, 易发生误操作的程度, 误操作容错能力};
  - 3)  $P_3=(P_{31}, P_{32})$ , 文档指标 = {培训, 帮助};
  - 4)  $I_1=(I_{11}, I_{12}, I_{13})$ , 信息交互指标 = {消息, 输入信息, 运行状态};
  - 5)  $I_2=(I_{21}, I_{22})$ , 可制定交互指标 = {界面内容, 操作方式}。
- d) 第四级因素层: 即具体指标集:
  - 1)  $P_{11}=\{P_{111}\}$ , 色彩 = {界面色彩对视觉的吸引力};
  - 2)  $P_{12}=\{P_{121}\}$ , 布局 = {布局的合理性};
  - 3)  $P_{13}=\{P_{131}\}$ , 一致性 = {一致性};
  - 4)  $P_{14}=\{P_{141}\}$ , 明显的功能 = {明显的功能};
  - 5)  $P_{21}=\{P_{211}, P_{212}\}$ , 操作复杂度 = {使用中默认值的可用性, 完成指定任务的步骤};
  - 6)  $P_{22}=\{P_{221}\}$ , 易发生误操作的程度 = {完成指定任务过程中误操作的次数};
  - 7)  $P_{23}=\{P_{231}, P_{232}, P_{233}\}$ , 误操作容错能力 = {错误的纠正, 发生错误的影响力, 可还原性};
  - 8)  $P_{31}=\{P_{311}\}$ , 培训 = {培训的有效性};
  - 9)  $P_{32}=\{P_{321}\}$ , 帮助 = {帮助机制的有效性}。
  - 10)  $I_{11}=\{I_{111}\}$ , 消息 = {使用中的消息的可理解性};
  - 11)  $I_{12}=\{I_{121}\}$ , 输入信息 = {输入的有效性检查};
  - 12)  $I_{13}=\{I_{131}\}$ , 运行状态 = {运行状态的易监控性};
  - 13)  $I_{21}=\{I_{211}, I_{212}\}$ , 界面内容 = {界面元素的易定制性, 界面布局的易定制性};
  - 14)  $I_{22}=\{I_{221}\}$ , 操作方式 = {快捷方式的易定制性}。

### A.3 网上评卷系统易用性评价集

评价集是对各层次评价指标的一种语言描述, 它是评审人对各评价指标所给出的评语的集合。网上评卷系统易用性评价指标体系的评语共分为四个等级。具体的评价集为:

$$V=(v_1, v_2, v_3, v_4)=(\text{优秀}, \text{良好}, \text{合格}, \text{不合格})$$

我们可以根据以上方法将网上评卷系统易用性各评价指标划分为四个等级, 而不同的人在执行不同的任务时所得出的指标值是完全不同的, 如果不考虑人和任务的因素去划分, 则评测结果将无法反应真实情况。我们已经根据工作内容不同将评卷教师分为了填空题组、作文题组和专家组, 并给出了各小组工作特点及人员特点。以下将为各评测指标划分评价集。

#### A.3.1 产品指标( $P$ )各级指标的评价集

##### A.3.1.1 外观指标( $P_1$ )

###### a) 色彩( $P_{11}$ )

界面色彩对视觉的吸引力( $P_{111}$ )

该指标的测量公式为：

$$X=A/B \qquad \qquad \qquad \dots\dots\dots (A.1)$$

式中：

A——在使用过评卷系统后,感觉界面配色方案舒适的用户人数；

B——使用过网上评卷系统的用户人数；

X——指标的度量值,0≤X≤1。

X越接近1,表明该系统的配色方案舒适度越好,越被用户所接受。

该指标是评测系统页面设计中色彩的搭配和运用是否合理,是否符合评卷教师的审美观点,同时是否易于缓解视觉疲劳。

该指标界面及功能设计是否符合业务流程,易于评卷教师理解功能。由于填空题组评卷教师年纪轻,适应能力强,故在评定该指标时应弱化该组评卷教师的感受,加重作文题组和专家组的意见,尤其是专家组。该指标各组的权值如下：

填空题:0.2,作文题组:0.3,专家组:0.5。

b) 布局(P<sub>12</sub>)

布局的合理性(P<sub>121</sub>)

该指标的测量公式为：

$$X=A/B \qquad \qquad \qquad \dots\dots\dots (A.2)$$

式中：

A——符合评卷工作要求及操作方便性要求的界面布局及元素数；

B——要求必须符合评卷工作要求及操作方便性要求的界面布局及元素数；

X——指标的度量值,0≤X≤1。

X越接近1,表明该评卷系统界面布局越符合评卷工作要求且操作方便性好。如表 A.1 所示。

表 A.1 布局的合理性(P<sub>121</sub>)的评价集

布局的合理性(P <sub>121</sub> )的评价集			
优秀	良好	合格	不合格
[1,0.75]	(0.75,0.5]	(0.5,0.375]	(0.375,0]

c) 一致性(P<sub>13</sub>)

一致性(P<sub>131</sub>)

该指标主要评测网上评卷系统的界面元素样式是否符合一致性要求。

测量公式为：

$$X=A/B \qquad \qquad \qquad \dots\dots\dots (A.3)$$

式中：

A——符合一致性要求的界面元素数；

B——要求必须符合一致性要求的界面元素的数；

X——指标的度量值,0≤X≤1。

X越接近1,表明该评卷系统界面元素一致性越好。如表 A.2 所示。

表 A.2 一致性( $P_{131}$ )的评价集

一致性( $P_{131}$ )的评价集			
优秀	良好	合格	不合格
[1,0.8]	(0.8,0.75]	(0.75,0.4]	(0.4,0]

d) 明显的功能( $P_{14}$ )

明显的功能( $P_{141}$ )

测量公式为：

$$X=A/B \qquad \qquad \qquad \dots\dots\dots (A.4)$$

式中：

A——在培训前,用户能够并可以根据界面提示进行操作的功能数；

B——要求在培训前,用户能够并可以根据界面提示进行操作的功能数；

X——指标的度量值,0≤X≤1。

X越接近1,表明该项指标越好。如表 A.3 所示。

表 A.3 明显的功能( $P_{141}$ )的评价集

明显的功能( $P_{141}$ )的评价集				
功能模块	优秀	良好	合格	不合格
评卷	[1,0.83]	(0.83,0.67]	(0.67,0.5]	(0.5,0.0]
质量监控	—	—	[1,0.5]	(0.5,0]

A.3.1.2 操作指标( $P_2$ )

a) 操作复杂度( $P_{21}$ )

使用中默认值的可用性( $P_{211}$ )

测量公式为：

$$X=A/B \qquad \qquad \qquad \dots\dots\dots (A.5)$$

式中：

A——评卷教师可自己选择的参数的功能数；

B——要求可供评卷教师自己选择参数的功能数；

X——指标的度量值,0≤X≤1。

X越接近1,表明该项指标越好。如表 A.4 所示。

表 A.4 使用中默认值的可用性( $P_{211}$ )的评价集

使用中默认值的可用性( $P_{211}$ )的评价集			
优秀	良好	合格	不合格
[1,0.8]	(0.8,0.75]	(0.75,0.4]	(0.4,0]

完成指定任务的步骤( $P_{212}$ )

测量公式为：

$$X=A/B \qquad \qquad \qquad \dots\dots\dots (A.6)$$

式中：  
A——符合完成所用的步骤数目的任务数；  
B——要求必须进行评测的任务数目；  
X——指标的度量值， $0 \leq X \leq 1$ 。  
X 越接近 1，表明该项指标越好。如表 A.5 所示。

表 A.5 完成指定任务的步骤(P<sub>212</sub>)的评价集

完成指定任务的步骤(P <sub>212</sub> )的评价集			
优秀	良好	合格	不合格
[1,0.8]	(0.8,0.75]	(0.75,0.4]	(0.4,0]

b) 易发生误操作的程度(P<sub>22</sub>)  
完成指定任务过程中误操作的次数(P<sub>221</sub>)  
测量公式为：  
$$X = A/B \quad \dots\dots\dots (A.7)$$

式中：  
A——用户在使用评卷系统完成某项任务时误操作的次数；  
B——用户在使用评卷系统完成某项任务时所用的步骤数目；  
X——指标的度量值， $X \geq 0$ 。  
X 越接近 0，表明该项指标越好。  
不同的任务，操作复杂度不同，出现错误的概率也各有不同，因此在考虑该指标时应考虑到各任务。  
如表 A.6 所示。

表 A.6 完成指定任务过程中误操作的次数(P<sub>221</sub>)的评价集

完成指定任务过程中误操作的次数(P <sub>221</sub> )的评价集				
任务名称	优秀	良好	合格	不合格
评卷	[0.3,0]	[0.3,0]	[0.3,0]	(+∞,0.3)
回评	(0.25,0]	(0.25,0]	(0.5,0.25]	(+∞,0.5)
问题卷的提交	[0.3,0]	[0.3,0]	[0.3,0]	(+∞,0.3)
设置雷同卷	[0.3,0]	[0.3,0]	[0.3,0]	(+∞,0.3)
查询培训情况	(0.25,0]	(0.25,0]	(0.5,0.25]	(+∞,0.5)
强制考核	(0.2,0]	(0.2,0]	(0.4,0.2]	(+∞,0.4)

c) 误操作容错能力(P<sub>23</sub>)  
错误的纠正(P<sub>231</sub>)  
测量公式为：  
$$X = A/B \quad \dots\dots\dots (A.8)$$

式中：  
A——用户在使用评卷系统完成某项任务时纠正错误的次数；  
B——用户在使用评卷系统完成某项任务时出现错误的次数；  
X——指标的度量值。

如表 A. 7 所示。

表 A. 7 错误的纠正( $P_{231}$ ) 的评价集

错误的纠正( $P_{231}$ ) 的评价集			
优秀	良好	合格	不合格
$[1,0.95)$	$[0.95,0.85)$	$[0.85,0.7)$	$[0.7,1]$

发生错误的影响力( $P_{232}$ )

测量公式为：

$$X=A/B \qquad \qquad \qquad \dots\dots\dots (A.9)$$

式中：

A——用户在使用评卷系统完成某项任务出现错误后，整个任务完成的时间；

B——用户在正常情况下，完成某项任务所用时间；

X——指标的度量值， $1\leq X\leq\infty$ 。

X 越接近 1，表明该项指标越好。如表 A. 8 所示。

表 A. 8 发生错误的影响力( $P_{232}$ ) 的评价集

发生错误的影响力( $P_{232}$ ) 的评价集			
优秀	良好	合格	不合格
$[1.1,1]$	$[1.2,1.1)$	$[1.3,1.2)$	$(+\infty,1.3)$

可还原性( $P_{233}$ )

用户在使用评卷系统完成某项功能时，界面状态是否能回到原状态。

测试方法同“布局的合理性”指标，测量公式为：

$$X=A/B \qquad \qquad \qquad \dots\dots\dots (A.10)$$

式中：

A——可以恢复到原状态的功能数；

B——要求可以恢复到原状态的功能数；

X——指标的度量值， $0\leq X\leq 1$ 。

X 越接近 1，表明该项指标越好。如表 A. 9 所示。

表 A. 9 可还原性( $P_{233}$ ) 的评价集

可还原性( $P_{233}$ ) 的评价集			
优秀	良好	合格	不合格
$[1,0.83)$	$[0.83,0.67)$	$[0.67,0.5)$	$[0.5,0]$

A. 3. 1. 3 文档指标( $P_3$ )

a) 培训( $P_{31}$ )

培训的有效性( $P_{311}$ )

测量公式为：

$$X=A/B \qquad \qquad \qquad \dots\dots\dots (A.11)$$



式中：  
A——用户在接受培训后可以理解和操作功能数；  
B——要求用户在接受培训后可以理解和操作的功能数；  
X——指标的度量值， $0 \leq X \leq 1$ 。  
X 越接近 1，表明该项指标越好。如表 A.10 所示。

表 A.10 培训的有效性( $P_{311}$ )的评价集

培训的有效性( $P_{311}$ )的评价集				
功能	优秀	良好	合格	不合格
评卷	[1,0.86)	[0.86,0.71)	[0.71,0.57)	[0.57,0.0]
质量监控	[1,0.75)	[0.75,0.5)	[0.75,0.5)	[0.5,0.0]

b) 帮助( $P_{32}$ )  
帮助机制的有效性( $P_{321}$ )

用户在使用评卷系统时,是否能够通过帮助系统来完成原先完成有困难的任务。该指标是反应系统提供的帮助系统对于用户完成任务是否是有效的。

$$X = A/B \quad \dots\dots\dots (A.12)$$

式中：  
A——用户在使用帮助后完成的任务数；  
B——用户在完成过程中因遇到困难而使用帮助系统的任务数；  
X——指标的度量值， $0 \leq X \leq 1$ 。  
X 越接近 1，表明该项指标越好。如表 A.11 所示。

表 A.11 帮助机制的有效性( $P_{321}$ )的评价集

帮助机制的有效性( $P_{321}$ )的评价集			
优秀	良好	合格	不合格
[1,0.8)	[0.8,0.65)	[0.65,0.4)	[0.4,0)

A.3.2 交互指标( $I$ )各级指标的评价集

A.3.2.1 信息交互指标( $I_1$ )

a) 消息( $I_{11}$ )  
使用中的消息的可理解性( $I_{111}$ )  
测量公式为：

$$X = A/B \quad \dots\dots\dots (A.13)$$

式中：  
A——在使用过评卷系统后,感觉系统提供的消息可以理解的人数；  
B——使用过网上评卷系统的用户人数；  
X——指标的度量值， $0 \leq X \leq 1$ 。  
X 越接近 1，表明该系统中的消息可理解性越好。

该指标的评测更依赖于评卷教师的主观感受,由于填空题组评卷教师年纪轻,适应能力强。另外,评卷功能相对于质量监控部分的功能较为简单,故在评定该指标时应弱化该组评卷教师的感受,加重作文题组和专家组的意见,尤其是专家组。该指标各组的权值如下：



填空题:0.1,作文题组:0.2,专家组:0.7。

b) 输入信息(I<sub>12</sub>)

输入的有效性检查(I<sub>121</sub>)

该指标是反应系统是否能够对分数、时间等关键输入信息进行有效性检查,以保证进入到系统数据的有效性。

测试方法同“布局的合理性”指标,故测量公式为:

$$X=A/B \dots\dots\dots (A.14)$$

式中:

A——能够对输入数据进行有效性检查的功能数;

B——要求对输入数据进行有效性检查的功能数;

X——指标的度量值,0≤X≤1。

X越接近1,表明该项指标越好。如表A.12所示。

表 A.12 输入的有效性检查(I<sub>121</sub>)的评价集

输入的有效性检查(I <sub>121</sub> )的评价集			
优秀	良好	合格	不合格
[1,0.83)	[0.83,0.67)	[0.67,0.5)	[0.5,0]

c) 运行状态(I<sub>13</sub>)

运行状态的易监控性(I<sub>131</sub>)

测量公式为:

$$X=A/B \dots\dots\dots (A.15)$$

式中:

A——能够向用户提供运行状态查看的功能数目;

B——要求向用户提供运行状态查看的功能数目;

X——指标的度量值,0≤X≤1。如表A.13所示。

表 A.13 运行状态的易监控性(I<sub>131</sub>)的评价集

运行状态的易监控性(I <sub>131</sub> )的评价集			
优秀	良好	合格	不合格
[1,0.8)	[0.8,0.6)	[0.6,0.5)	[0.4,0]

A.3.2.2 可定制交互指标(I<sub>2</sub>)

a) 界面内容(I<sub>21</sub>)

界面元素的易定制性(I<sub>211</sub>)

测试方法同“布局的合理性”指标,故测量公式为:

$$X=A/B \dots\dots\dots (A.16)$$

式中:

A——用户可以改变的界面元素数;

B——要求用户可以改变的界面元素数;

X——指标的度量值,0≤X≤1。

X越接近1,表明该项指标越好。如表A.14所示。

表 A.14 界面元素的易定制性(I<sub>211</sub>)的评价集

界面元素的易定制性(I <sub>211</sub> )的评价集			
优秀	良好	合格	不合格
[1,0.75)	[0.75,0.5)	[0.5,0.2)	[0.2,0]

界面布局的易定制性(I<sub>212</sub>)

测量公式为：

$$X=A/B \qquad \qquad \qquad \dots\dots\dots (A.17)$$

式中：

- A——用户可以改变的界面布局元素数；
- B——要求用户可以改变的界面元素数；
- X——指标的度量值,0≤X≤1。
- X 越接近 1,表明该项指标越好。
- 用户可以定制的界面布局如表 A.15 所示。

表 A.15 界面布局的易定制性(I<sub>212</sub>)的评价集

界面布局的易定制性(I <sub>212</sub> )的评价集			
优秀	良好	合格	不合格
[1,0.67)	[1,0.67)	[1,0.67)	[0.67,0]

b) 操作方式(I<sub>22</sub>)

快捷方式的易定制性(I<sub>221</sub>)

测量公式为：

$$X=A/B \qquad \qquad \qquad \dots\dots\dots (A.18)$$

式中：

- A——用户可以自定义快捷键的功能数；
- B——要求用户可以自定义快捷键的功能数；
- X——指标的度量值,0≤X≤1。
- X 越接近 1,表明该项指标越好。如表 A.16 所示。

表 A.16 快捷方式的易定制性(I<sub>221</sub>)的评价集

快捷方式的易定制性(I <sub>221</sub> )的评价集			
优秀	良好	合格	不合格
[1,0.8)	[0.8,0.6)	[0.6,0.4)	[0.4,0]

A.4 网上评卷系统易用性各评测指标的权重

A.4.1 网上评卷系统易用性各评测指标权重计算方法简述

权重实际上是指人们在评判事物时,依次着重于哪些指标。用于确定“权重”的方法,主要有两类。一类是专家咨询法。即向一批有经验的专家发调查表,让他们对每一个指标在综合评价时应占的“权重”分别填写出自己的意见,回收调查表后,进行统计分析,根据多数专家的意见来确定“权重”。另一类是数理统计法,即通过测试一批数据后,用 R 型因子分析等多因素统计方法进行计算,根据计算的结果

来确定“权重”。两种方法各有优势及不足,本附录采用的一种将两种方法相结合的方法—指标序号法,即向专家咨询的调查表。在填表说明中规定:根据专家自己的看法,认为哪个指标在综合评价中最重要,它的序号就填为 1,哪个次之,就填为 2;如果认为某几个指标的重要程度一样,就可以填相同的序号;假如认为某指标很不重要,在综合评价时可以不要,则序号不填。以下将结合该方法来讨论如何网上评卷系统易用性各评价指标的权重。

a) 专家填写调查问卷。选择 20 位网上评卷专家,这些专家包括评卷的管理者,评卷教师,评卷组长及网上评卷系统开发人员。网上评卷系统易用性评价指标分为产品指标(P)和交互指标(I)两大类,我们可以规定这两类指标的权重各为 0.5,并且请专家分别对这两类指标下的各级指标进行排序。各指标重要程度的排序是分类进行的,不仅要对评价指标结构中的第一层指标进行排序,还要对各指标的子指标分别进行排序。调查问卷如表 A.17 所示。

表 A.17 网上评卷系统易用性评价指标重要程度调查表(产品指标)

指标类别	重要程度序号	指标集	重要程度序号	具体指标	重要程度序号
外观指标		色彩		界面色彩对视觉的吸引性	
		布局		布局的合理性	
		一致性		一致性	
		明显的功能		明显的功能	
操作指标		操作复杂度		使用中默认值的可用性	
				完成指定任务的步骤	
		易发生误操作的程度		完成指定任务过程中误操作的次数	
		误操作容错能力		错误的纠正	
				发生错误的影响力	
				可还原性	
文档指标		培训		培训的有效性	
		帮助		帮助机制的有效性	

请首先对类别的重要程度进行排序,序号为 1~3;之后请对各类别下属的指标类别分别排序。调查表如表 A.18 所示。

表 A.18 网上评卷系统易用性评价指标重要程度调查表(交互指标)

指标类别	重要程度序号	指标集	重要程度序号	具体指标	重要程度序号
信息交互指标		消息		使用中的消息的可理解性	
		输入信息		输入的有效性检查	
		运行状态		运行状态的易监控性	
可定制交互指标		界面内容		界面元素的易定制性	
				界面布局的易定制性	
		操作方式		快捷方式的易定制性	

请首先对类别的重要程度进行排序,序号为 1~2;之后请对各类别下属的指标类别分别排序。

b) 统计计算权重值。调查表回收后,进行统计计算时,每一大类要画一张“权重”计算表,表内的序号列数是根据各类的指标数而定的,如吸引力指标有 2 个指标。首先,把回收的调查表内所填的各指标序号,像填频数分布表一样,用画“正”字的办法,逐张进行登录,把全部调查表登录完毕后,再逐格统计频数。分别计算出各指标各频数之和,其占有所有指标频数总和的比例,则为各指标的“权重”。计算公式如下:

$$a_i = \frac{F = \sum [(n+1-i)f_i]}{n \sum i} \dots\dots\dots (A.19)$$

式中:  
n —— 指标数;  
i —— 序号;  
f<sub>i</sub> —— 序号 i 内的频数。

A.4.2 网上评卷系统易用性各评测指标权重值的计算过程

A.4.2.1 产品指标(P)各级指标权重值的计算过程

a) 第二层指标的权重值  
如表 A.19 所示。

表 A.19 网上评卷系统易用性评测产品指标(P)“权重计算表”

指标	频数			权重	
	1	2	3	F	a <sub>i</sub>
外观指标	8	7	5	8×3+7×2+5×1=43	0.358
操作指标	9	9	2	9×3+9×2+2×1=47	0.392
文档指标	3	4	13	3×3+4×2+13×1=30	0.250
Σ				120	1

指标数 n=20。  
根据表 A.19 的计算结果,各子指标的权重值为:  
a=(a<sub>1</sub>,a<sub>2</sub>,a<sub>3</sub>)=(0.358,0.392,0.250)  
a<sub>1</sub>:外观指标;a<sub>2</sub>:操作指标;a<sub>3</sub>:文档指标。

b) 第三层指标的权重值  
如表 A.20 所示。

表 A.20 网上评卷系统易用性外观指标(P<sub>1</sub>)指标“权重计算表”

指标	频数				权重	
	1	2	3	4	F	a <sub>i</sub>
色彩	2	3	6	9	2×4+3×3+6×2+9×1=38	0.190
布局	8	7	4	1	8×4+7×3+4×2+1×1=62	0.310
一致性	3	2	7	8	3×4+2×3+7×2+8×1=40	0.200
明显的功能	7	8	3	2	7×4+8×3+3×2+2×1=60	0.300
Σ					200	1

指标数 n=20。

根据表 A. 20 的计算结果,各子指标的权重值为:  
 $a_1=(a_{11},a_{12},a_{13},a_{14})=(0.190,0.310,0.200,0.300)$   
 $a_{11}$ :色彩; $a_{12}$ :布局; $a_{13}$ :一致性; $a_{14}$ :明显的功能。  
如表 A. 21 所示。

表 A. 21 网上评卷系统易用性操作指标( $P_2$ )指标“权重计算表”

指标	频数			权重	
	1	2	3	$F$	$a_i$
操作复杂度	9	8	3	$9\times3+8\times2+3\times1=46$	0.383
易发生误操作的程度	3	5	12	$3\times3+5\times2+12\times1=31$	0.258
误操作的容错能力	8	7	5	$8\times3+7\times2+5\times1=43$	0.358
$\Sigma$				120	1

指标数  $n=20$ 。  
根据表 A. 21 的计算结果,各子指标的权重值为:  
 $a_2=(a_{21},a_{22},a_{23})=(0.383,0.258,0.358)$   
 $a_{21}$ :操作复杂度; $a_{22}$ :易发生误操作的程度; $a_{23}$ :误操作的容错能力。  
如表 A. 22 所示。

表 A. 22 网上评卷系统易用性文档指标( $P_3$ )指标“权重计算表”

指标	频数		权重	
	1	2	$F$	$a_i$
培训	15	5	$15\times2+5\times1=35$	0.583
帮助	5	15	$5\times2+15\times1=25$	0.417
$\Sigma$			60	1

指标数  $n=20$ 。  
根据表 A. 22 的计算结果,各子指标的权重值为:  
 $a_3=(a_{31},a_{32})=(0.583,0.417)$   
 $a_{31}$ :培训; $a_{32}$ :帮助。  
c) 具体指标的权重值  
根据网上评卷易用性评价指标体系结构中,我们可以看出在具体指标层中对应  $P_{11}$ 、 $P_{12}$ 、 $P_{13}$ 、 $P_{14}$ 、 $P_{22}$ 、 $P_{31}$ 、 $P_{32}$  指标的具体指标只有一个,因此我们只需计算  $P_{21}$  和  $P_{23}$  所对应的具体指标的权重值即可。  
如表 A. 23、表 A. 24 所示。

表 A. 23 网上评卷系统易用性操作复杂度( $P_{21}$ )指标“权重计算表”

指标	频数		权重	
	1	2	$F$	$a_i$
使用中默认值的可用性	3	17	$3\times2+17\times1=23$	0.383
完成制定任务的步骤	17	3	$17\times2+3\times1=37$	0.617
$\Sigma$			60	1



指标数  $n=20$ 。  
根据表 A. 23 的计算结果,各子指标的权重值为:  
 $a_{21}=(a_{211},a_{212})=(0.383,0.617)$   
 $a_{211}$ :使用中默认值的可用性; $a_{212}$ :完成指定任务的步骤。

表 A. 24 网上评卷系统易用性误操作的容错能力( $P_{23}$ )指标“权重计算表”

指标	频数			权重	
	1	2	3	$F$	$a_i$
错误的纠正	4	9	7	$4\times3+9\times2+7\times1=37$	0.308
发生错误后完成任务的时间	13	3	4	$13\times3+3\times2+4\times1=49$	0.408
可还原性	3	8	9	$3\times3+8\times2+9\times1=34$	0.283
$\Sigma$				120	1

指标数  $n=20$ 。  
根据表 A. 24 的计算结果,各子指标的权重值为:  
 $a_{23}=(a_{231},a_{232},a_{233})=(0.308,0.408,0.283)$   
 $a_{231}$ :错误的纠正; $a_{232}$ :发生错误的影响力; $a_{233}$ :可还原性。

A. 4. 2. 2 交互指标( $I$ )各级指标权重的计算过程

a) 第二层指标的权重值  
如表 A. 25 所示。

表 A. 25 网上评卷系统易用性评测交互指标( $I$ )“权重计算表”

指标	频数		权重	
	1	2	$F$	$a_i$
信息交互指标	16	4	$16\times2+4\times1=36$	0.600
可定制指标	4	16	$4\times2+16\times1=24$	0.400
$\Sigma$			60	1

指标数  $n=20$ 。  
根据表 A. 25 的计算结果,各子指标的权重值为:  
 $a=(a_1,a_2)=(0.600,0.400)$   
 $a_1$ :信息交互指标; $a_2$ :可定制交互指标。  
b) 第三层指标的权重值  
如表 A. 26 所示。

表 A. 26 网上评卷系统易用性评测信息交互指标( $I_1$ )“权重计算表”

指标	频数			权重	
	1	2	3	$F$	$a_i$
消息	6	5	9	$6\times3+5\times2+9\times1=37$	0.308
输入信息	8	10	2	$8\times3+10\times2+2\times1=46$	0.383
运行状态	6	5	9	$6\times3+5\times2+9\times1=37$	0.308
$\Sigma$				120	1



指标数  $n=20$ 。  
根据表 A. 26 的计算结果,各子指标的权重值为:  
 $a_1=(a_{11},a_{12},a_{13})=(0.308,0.383,0.308)$   
 $a_{11}$ :消息; $a_{12}$ :输入信息; $a_{13}$ :运行状态。  
如表 A. 27 所示。

表 A. 27 网上评卷系统易用性评测可定制交互指标( $I_2$ )“权重计算表”

指标	频数		权重	
	1	2	$F$	$a_i$
界面内容	11	9	$11 \times 2 + 9 \times 1 = 31$	0.517
操作方式	9	11	$9 \times 2 + 11 \times 1 = 29$	0.483
$\Sigma$			60	1

指标数  $n=20$ 。  
根据表 A. 27 的计算结果,各子指标的权重值为:  
 $a_2=(a_{21},a_{22})=(0.517,0.483)$   
 $a_{21}$ :界面内容; $a_{22}$ :操作方式。  
c) 具体指标值的权重值

同产品指标,我们可以看出在具体指标层中对应  $I_{11}, I_{12}, I_{13}, I_{22}$  指标的具体指标只有一个,因此我们只需计算  $I_{21}$  所对应的具体指标的权重值即可。如表 A. 28 所示。

表 A. 28 网上评卷系统易用性评测可定制指标( $I_{21}$ )“权重计算表”

指标	频数		权重	
	1	2	$F$	$a_i$
界面元素的易定制性	17	3	$17 \times 2 + 3 \times 1 = 37$	0.617
界面布局的易定制性	3	17	$3 \times 2 + 17 \times 1 = 23$	0.383
$\Sigma$			60	1

指标数  $n=20$ 。  
根据表 A. 28 的计算结果,各子指标的权重值为:  
 $a_{21}=(a_{211},a_{212})=(0.617,0.383)$   
 $a_{21}$ :界面元素的易定制性; $a_{22}$ :界面布局的易定制性。

A. 5 构造网上评卷系统易用性模糊判断矩阵

构造模糊判断矩阵首先对某个指标集中的单个指标  $u_i(i=1,2,3,\cdots,m)$  做单因素判断,从指标因素  $u_i$  对选择等级  $v_j(j=1,2,3,\cdots,n)$  的隶属度为  $r_{ij}$ ,这样就得出第  $i$  个指标  $u_i$  的单指标判断集:

$r_i=(r_{i1},r_{i2},\cdots,r_{in})$

这样  $m$  个指标的评价集就构造出一个总的评价矩阵  $R$ 。即每一个被评价对象确定了从  $U$  到  $V$  的模糊关系  $R$ ,它是一个矩阵:

$$R=(r_{ij})_{m \times n} = \begin{bmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mn} \end{bmatrix} (i=1,2,\cdots,m;j=1,2,\cdots,n) \cdots \cdots \cdots (A.20)$$

式中  $r_{ij}$  表示从指标  $u_i$  着眼,该评判对象能被评为  $v_j$  的隶属度。具体地说, $r_{ij}$  表示第  $i$  个指标  $u_i$  在第  $j$  个评语  $v_j$  上的频率分布。

下面将结合以上概念及网上评卷系统易用性评测指标来分层构造网上评卷系统易用性模糊判断矩阵:

评价集为  $V=(v_1, v_2, v_3, v_4)=(\text{优秀}, \text{良好}, \text{合格}, \text{不合格})$

a) 总目标层

其指标集为:  $U=(P, I)$

则模糊判断矩阵为:

$$R=(r_{ij})_{m \times n}=\begin{bmatrix} r_{11} & \cdots & r_{14} \\ r_{21} & \cdots & r_{24} \end{bmatrix} (i=2; j=4) \cdots \cdots (A.21)$$

b) 第二层中产品指标

其指标集为:  $P=(P_1, P_2, P_3)$

则模糊判断矩阵为:

$$R'_1=(r'_{ij})_{m \times n}=\begin{bmatrix} r'_{11} & \cdots & r'_{14} \\ \vdots & \ddots & \vdots \\ r'_{31} & \cdots & r'_{34} \end{bmatrix} (i=3; j=4) \cdots \cdots (A.22)$$

$P_1$  的子指标集为:  $P_1=(P_{11}, P_{12}, P_{13}, P_{14})$

则模糊判断矩阵为:

$$R'_1=(r'_{ij})_{m \times n}=\begin{bmatrix} r'_{11} & \cdots & r'_{14} \\ \vdots & \ddots & \vdots \\ r'_{41} & \cdots & r'_{44} \end{bmatrix} (i=4; j=4) \cdots \cdots (A.23)$$

$P_2$  的子指标集为:  $P_2=(P_{21}, P_{22}, P_{23})$

则模糊判断矩阵为:

$$R'_2=(r'_{ij})_{m \times n}=\begin{bmatrix} r'_{11} & \cdots & r'_{14} \\ \vdots & \ddots & \vdots \\ r'_{31} & \cdots & r'_{34} \end{bmatrix} (i=3; j=4) \cdots \cdots (A.24)$$

$P_3$  的子指标集为:  $P_3=(P_{31}, P_{32})$

则模糊判断矩阵为:

$$R'_3=(r'_{ij})_{m \times n}=\begin{bmatrix} r'_{11} & \cdots & r'_{14} \\ r'_{21} & \cdots & r'_{24} \end{bmatrix} (i=2; j=4) \cdots \cdots (A.25)$$

c) 第二层中交互指标

其指标集为  $I=(I_1, I_2)$

则模糊判断矩阵为:

$$R''_1=(r''_{ij})_{m \times n}=\begin{bmatrix} r''_{11} & \cdots & r''_{14} \\ r''_{21} & \cdots & r''_{24} \end{bmatrix} (i=2; j=4) \cdots \cdots (A.26)$$

$I_1$  的子指标集为:  $I_1=(I_{11}, I_{12}, I_{13})$

则模糊判断矩阵为:

$$R''_2=(r''_{ij})_{m \times n}=\begin{bmatrix} r''_{11} & \cdots & r''_{14} \\ \vdots & \ddots & \vdots \\ r''_{31} & \cdots & r''_{34} \end{bmatrix} (i=3; j=4) \cdots \cdots (A.27)$$

$I_2$  的子指标集为:  $I_2=(I_{21}, I_{22})$

则模糊判断矩阵为:

$$R''_3 = (r'_{ij})_{m \times n} = \begin{bmatrix} r'_{11} & \cdots & r'_{14} \\ r'_{21} & \cdots & r'_{24} \end{bmatrix} (i=2; j=4) \cdots \cdots (A.28)$$

确定模糊判断矩阵的值。从定义中我们知道  $r_{ij}$  表示第  $i$  个指标  $u_i$  在第  $j$  个评语  $v_j$  上的频率分布, 在网上评卷系统易用性评测过程中, 我们可以将  $r_{ij}$  为网上评卷系统易用性某评测指标为评价集中各值 (优秀, 良好, 合格, 不合格) 的概率各是多少。我们可以将 A.3 给出的各评测指标值划分成优秀, 良好, 合格, 不合格四个取值区间, 具体划分方法请见 A.3 中各表。通过对测试用户所填写调查问卷的分析, 得出测试结果在优秀, 良好, 合格, 不合格四个区间内的人数占总测试人数的比例。该比例则是  $r_{ij}$ 。

$$r_{ij} = \frac{\text{评价指标 } u_i \text{ 中测试结果为 } v_j \text{ 的人数}}{\text{参与测试的总人数 } n} \cdots \cdots (A.29)$$

例如, 通过对调查问卷的分析, 对评测指标“明显的功能” $u_{11}$  测试结果为优秀的人数为 10, 参与测试的人员为 100 人, 则该指标在优秀上的频率分布为 0.1。

## A.6 综合评价

模糊矩阵  $R$  中不同的行反应了某个被评价事物中不同的单因素 (单指标) 来看对各等级模糊子集  $r_{ij}$  的隶属程度。用模糊权向量  $A$  将不同的行进行综合, 就可得到该被评事物从总体上来看对各等级模糊子集的隶属程度, 即模糊综合评价结果向量。

引入评价集  $V$  上的一个模糊子集  $B$ , 称模糊评价, 又称决策集。  $B = (b_1, b_2, \dots, b_n)$ 。  $B$  是对每个被评判对象综合状况分等级的程度描述, 通常可以采用最大隶属度法则对其处理, 得到最终评判结果。例如在对某网上评卷系统易用性进行综合评价时, 评价集  $v = (v_1, v_2, v_3, v_4) = (\text{优秀}, \text{良好}, \text{及格}, \text{不及格})$ , 得出决策集  $B = (0.3, 0.2, 0.4, 0.1)$ , 根据最大隶属度法, 则该网上评卷系统易用性的评测结果为“及格”(0.4)。

网上评卷系统评价体系分为三层。在进行综合评价时首先对最低层目标层各指标评价, 通过对调查结果的整理、统计, 即得到单指标模糊判断矩阵。分别将各单指标模糊判断矩阵与相应的模糊权向量进行模糊运算, 得出决策集。首先, 计算产品指标第二层的决策集

$$B'_1 = a_1 \circ R'_1 = (a_{11}, a_{12}, a_{13}, a_{14}) \circ \begin{bmatrix} r'_{11} & \cdots & r'_{14} \\ \vdots & \ddots & \vdots \\ r'_{41} & \cdots & r'_{44} \end{bmatrix} = (B'_{11}, B'_{12}, B'_{13}, B'_{14}) \cdots \cdots (A.30)$$

$$B'_2 = a_2 \circ R'_2 = (a_{21}, a_{22}, a_{23}) \circ \begin{bmatrix} r'_{11} & \cdots & r'_{14} \\ \vdots & \ddots & \vdots \\ r'_{31} & \cdots & r'_{34} \end{bmatrix} = (B'_{21}, B'_{22}, B'_{23}, B'_{24}) \cdots \cdots (A.31)$$

$$B'_3 = a_3 \circ R'_3 = (a_{31}, a_{32}) \circ \begin{bmatrix} r'_{11} & \cdots & r'_{14} \\ r'_{21} & \cdots & r'_{24} \end{bmatrix} = (B'_{31}, B'_{32}, B'_{33}, B'_{34}) \cdots \cdots (A.32)$$

根据以上得出的各子目标层指标模糊评价判断矩阵, 得出产品指标的决策集模糊判断矩阵为:

$$R' = \begin{bmatrix} B'_1 \\ B'_2 \\ B'_3 \end{bmatrix} \cdots \cdots (A.33)$$

产品指标的决策集为:

$$B' = a \circ R' = (a_1, a_2, a_3) \circ \begin{bmatrix} B'_1 \\ B'_2 \\ B'_3 \end{bmatrix} = (b'_1, b'_2, b'_3, b'_4) \cdots \cdots (A.34)$$

同理得出交互指标的决策集为

$$B' = c^{\circ}R'' = (c_1, c_2)^{\circ} \begin{bmatrix} B''_1 \\ B''_2 \end{bmatrix} = (b'_1, b'_2, b'_3, b'_4) \dots\dots\dots (A. 35)$$

根据以上结果,得出总目标  $U$  的决策集

$$B = A^{\circ}R = (a, c)^{\circ} \begin{bmatrix} B' \\ B'' \end{bmatrix} = (b_1, b_2, b_3, b_4) \dots\dots\dots (A. 36)$$

大隶属度法,得出被评测网上评卷系统易用性的评测结果。

## 参 考 文 献

- [1] GB/T 5271.1 信息技术 词汇 第1部分:基本术语(GB/T 5271.1—2000,eqv ISO/IEC 2382-1:1993)
  - [2] GB/T 5271.20 信息技术 词汇 20部分:系统开发(GB/T 5271.20—1994,eqv ISO/IEC 2382-20:1990)
  - [3] GB/T 8566—2007 信息技术 软件生存周期过程(ISO/IEC 12207:1995,IDT)
  - [4] GB/T 18905.2—2002 软件工程 产品评价 第2部分:策划和管理(ISO/IEC 14598-2:2000,IDT)
  - [5] GB/T 18905.3—2002 软件工程 产品评价 第3部分:开发者用的过程(ISO/IEC 14598-3:2000,IDT)
  - [6] GB/T 18905.4—2002 软件工程 产品评价 第4部分:需方用的过程(ISO/IEC 14598-4:1999,IDT)
  - [7] GB/T 18905.5—2002 软件工程 产品评价 第5部分:评价者用的过程(ISO/IEC 14598-5:1998,IDT)
  - [8] GB/T 18905.6—2002 软件工程 产品评价 第6部分:评价模块的文档编制(ISO/IEC 14598-6:2001,IDT)
  - [9] ISO 9241-10:1996 Ergonomic requirements for office work with visual display terminals (VDTs)—Part 10:Dialogue principles
  - [10] ISO 9241-11:1998 Ergonomic requirements for office work with visual display terminals (VDTs)—Part 11:Guidance on usability
  - [11] ISO 13407:1999 Human-centred design processes for interactive systems
  - [12] ISO/IEC TR 15504 Information technology—Process assessment
-

中 华 人 民 共 和 国  
国 家 标 准  
系统与软件易用性 第2部分:度量方法  
GB/T 29836.2—2013

\*

中国标准出版社出版发行  
北京市朝阳区和平里西街甲2号(100013)  
北京市西城区三里河北街16号(100045)  
网址 [www.spc.net.cn](http://www.spc.net.cn)  
总编室:(010)64275323 发行中心:(010)51780235  
读者服务部:(010)68523946  
中国标准出版社秦皇岛印刷厂印刷  
各地新华书店经销

\*

开本 880×1230 1/16 印张 2 字数 56 千字  
2014 年 1 月第一版 2014 年 1 月第一次印刷

\*

书号: 155066·1-48017 定价 30.00 元

如有印装差错 由本社发行中心调换  
版权专有 侵权必究  
举报电话:(010)68510107



GB/T 29836.2-2013