



(12) 发明专利申请

(10) 申请公布号 CN 104318093 A

(43) 申请公布日 2015. 01. 28

(21) 申请号 201410544919. 7

(22) 申请日 2014. 10. 15

(71) 申请人 昆明理工大学

地址 650093 云南省昆明市五华区学府路
253 号

(72) 发明人 邵玉斌 罗胜 龙华 杜庆治

(51) Int. Cl.

G06F 19/00 (2011. 01)

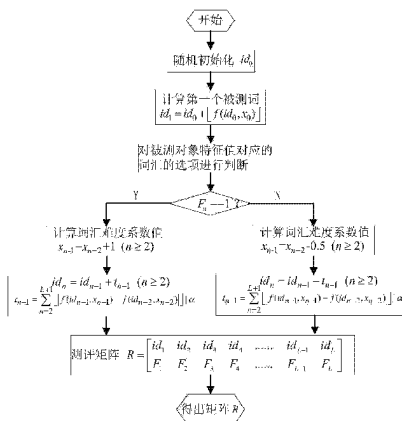
权利要求书2页 说明书9页 附图3页

(54) 发明名称

一种基于马尔科夫的词汇量测试与评估方法

(57) 摘要

本发明涉及一种基于马尔科夫的词汇量测试与评估方法,属于语言学习领域。本发明首先确定所有词汇的属性值及进行难度级别划分;接着随机从第 1 个难度级别的词汇中选取一个属性值作为初始特征值并确定词汇难度系数;然后计算第一个测试词汇的对应的特征值并对其进行判断;再接着计算特征值所对应的难度系数值,同时根据状态值确定下一个特征值;最后根据测试评价对象矩阵处理得出受试者是否掌握测试结果并得出受试者的词汇测评域估计值。本发明克服目前测试词汇中测试词没有关联性的缺陷,也更准确的反应被测者的真实水平词汇水平。



1. 一种基于马尔科夫的词汇量测试与评估方法,其特征在于:所述方法的具体步骤如下:

Step1、确定所有词汇的属性值为 $1, \dots, i (i \in \mathbb{N}^*)$, 并以每 1500 个词汇划分为 1 个难度级别;其中,最后小于 1500 的词汇归于最后一个难度级别;

Step2、随机从第 1 个难度级别的词汇中选取一个属性值作为初始特征值 id_0 , 并确定词汇难度系数为 x_0 ;其中, $x_0 = 1$ 表示初始的难度系数值;

Step3、计算第一个测试词汇对应的特征值: $id_1 = id_0 + [f(id_0, x_0)]$; 其中, $f(id_0, x_0)$ 为初始特征值生成函数;

Step4、对被测对象的特征值 id_1 对应的词汇的选项进行判断:

如果被测对象的特征值 id_1 对应的词汇的选项正确,则定义标记 $F_1 = 1$;

如果被测对象的特征值 id_1 对应的词汇的选项错误,则定义标记 $F_1 = 0$;

Step5、计算 id_{n-1} 所对应的词汇难度系数 x_{n-1} 的取值:

如果 $F_{n-1} = 1$, 则 $x_{n-1} = x_{n-2} + 1$;

如果 $F_{n-1} = 0$, 则 $x_{n-1} = x_{n-2} - 0.5$;

其中, $n \geq 2$;

Step6、根据 F_{n-1} 的取值,确定下一个被测对象的特征值 id_n :

如果 $F_{n-1} = 1$, 则 $id_n = id_{n-1} + t_{n-1}$;

如果 $F_{n-1} = 0$, 则 $id_n = id_{n-1} - t_{n-1}$;

其中, $t_{n-1} = \sum_{n=2}^{L+1} [|f(id_{n-1}, x_{n-1}) - f(id_{n-2}, x_{n-2})|] \alpha$, L 表示测试的总个数且 $L \geq 1$; x_{n-1}

为根据词汇特征值取 id_{n-1} 时所对应的难度系数值; x_{n-2} 为根据词汇特征值取 id_{n-2} 时所对应的难度系数值; $f(id_{n-1}, x_{n-1})$ 表示当难度系数值为 x_{n-1} 时 id_{n-1} 所对应的特征值生成函数; $f(id_{n-2}, x_{n-2})$ 表示当难度系数值为 x_{n-2} 时 id_{n-2} 所对应的特征值生成函数; α 的符号由 id_{n-1} 对应标记 F_{n-1} 的取值决定:如果 $F_{n-1} = 1$, 则符号为正,否则为负;

Step7、根据受试者对已经测试过的对象的特征值及其状态值得出一个 $2 \times L$ 测试结果矩阵 R :

$$R = \begin{bmatrix} id_1 & id_2 & id_3 & id_4 & \dots & id_{L-1} & id_L \\ F_1 & F_2 & F_3 & F_4 & \dots & F_{L-1} & F_L \end{bmatrix}$$

其中, L 表示测试的总个数且 $L \geq 1$;

Step8、将测试结果矩阵 R 进行处理得出受试者是否掌握测试结果:

如果特征值 id_l 对应的 F_l 为 1 时,则认为在 id_n 左右邻域中的词汇受试者都能掌握;

如果特征值 id_l 对应的 F_l 为 0 时,则认为在 id_n 左右邻域中的词汇受试者没有掌握;

Step9、根据测试结果,得出受试者的词汇测评域估计值为:

$$result = \sum_{k=1}^N 2 \times \psi_k(id_k) - \sum_{j=1}^M 2 \times \varphi_j(id_j)$$

其中, N 表示测试判断对的总个数, M 表示测试判断错的总个数; $\psi_k(id_k)$ 表示被测对

象特征值 id_k 对应的词汇的选项正确时的左右邻域的取值, $\varphi_j(id_j)$ 表示被测对象特征值 id_j 对应的词汇的选项错误时的左右邻域的取值。

2. 根据权利要求 1 所述的基于马尔科夫的词汇量测试与评估方法,其特征在于:所述步骤 Step6 能替换为:

如果 $F_{n-1} = 1$, 则 $id_n = id_{n-1} + t_{n-1}$;

如果 $F_{n-1} = 0$, 则 $id_n = id_{n-1} - t_{n-1}$;

其中, $t_{n-1} = \sum_{n=2}^{L+1} \zeta_{n-1} [|f(id_{n-1}, x_{n-1}) - f(id_{n-2}, x_{n-2})|] \alpha$, L 表示测试的总个数且 $L \geq 1$;;

ζ_{n-1} 表示 $[|f(id_{n-1}, x_{n-1}) - f(id_{n-2}, x_{n-2})|]$ 的加权值, $\zeta_{n-1} > \zeta_{n-2} > \zeta_{n-3} > \dots > \zeta_2 > \zeta_1$ 且 $\zeta_{n-1} + \zeta_{n-2} + \dots + \zeta_2 + \zeta_1 = 1$; 其中, α 的符号由 id_{n-1} 对应标记 F_{n-1} 的取值决定: 如果 $F_{n-1} = 1$, 则符号为正, 否则为负。

3. 根据权利要求 1 或 2 所述的基于马尔科夫的词汇量测试与评估方法,其特征在于:

所述特征值生成函数 $f(id_{n-2}, x_{n-2}) = \frac{id_{n-2}}{1 + Ae^{-x_{n-2}}}$; 其中, A 为函数特征系数, $n \geq 2$ 。

一种基于马尔科夫的词汇量测试与评估方法

技术领域

[0001] 本发明涉及一种基于马尔科夫的词汇量测试与评估方法,属于语言学习领域。

背景技术

[0002] 随着社会的发展,很多的测试评价都是智能的,尤其在语言词汇量的测试中。采取什么样的方法才能更准确的测试出受试者的语言词汇量水平,怎样设计更为合理的计算方法就成了关键因数,一个合理的测试计算方法能准备的反应出一个受试者的真正水平。尤其在英语词汇的测试中怎样通过一定量的测试结果来预测出受试者的英语词汇量水平。在一种随机的测试中,怎样把受试者已经测试过的样本词汇结果和将要测试的词汇关联起来。怎样在已测试过的样本词汇结果的基础上对将要测试词汇的难度进行随机且适当的调整。这是一般标准化考试或测试中很难做到的。

[0003] 为了克服上述中智能化标准测试中的评价关联性及其难度系数的关联性,本发明提出一种语言词汇量测试与评估方法。随机测试的下一个词汇与已测词汇的马儿可夫性质,也可以是多维的马儿可夫性质,及测试词汇的难度也具有马儿可夫性质。

发明内容

[0004] 本发明提供了一种基于马尔科夫的词汇量测试与评估方法,以用于克服智能化词汇标准测试中的测试词汇之间没关联性的缺陷,以及测试词汇的难度系数不可调整的缺陷。除了克服其缺陷以外,测试词汇之间存在着严格的关联性,而且关联程度和将要测试的词汇的距离有关,即从该测试词汇起直到以前 n 个词的测试状态的影响,离该词远的其测试状态对现在的影响就越小,这样使得测试词汇之间的关联程度更加的紧密。使得测试的所有词汇之间紧密关联。

[0005] 本发明的技术方案是:一种基于马尔科夫的词汇量测试与评估方法,所述方法的具体步骤如下:

[0006] Step1、确定所有词汇的属性值为 $1 \dots i (i \in \mathbb{N}^*)$,并以每 1500 个词汇划分为 1 个难度级别;其中,最后小于 1500 的词汇归于最后一个难度级别;

[0007] Step2、随机从第 1 个难度级别的词汇中选取一个属性值作为初始特征值 id_0 ,并确定词汇难度系数为 x_0 ;其中, $x_0 = 1$ 表示初始的难度系数值;

[0008] Step3、计算第一个测试词汇对应的特征值: $id_1 = id_0 + [f(id_0, x_0)]$;其中, $f(id_0, x_0)$ 为初始特征值生成函数;

[0009] Step4、对被测对象的特征值 id_1 对应的词汇的选项进行判断:

[0010] 如果被测对象的特征值 id_1 对应的词汇的选项正确,则定义标记 $F_1 = 1$;

[0011] 如果被测对象的特征值 id_1 对应的词汇的选项错误,则定义标记 $F_1 = 0$;

[0012] Step5、计算 id_{n-1} 所对应的词汇难度系数 x_{n-1} 的取值:

[0013] 如果 $F_{n-1} = 1$,则 $x_{n-1} = x_{n-2} + 1$;

[0014] 如果 $F_{n-1} = 0$,则 $x_{n-1} = x_{n-2} - 0.5$;

[0015] 其中, $n \geq 2$;

[0016] Step6、根据 F_{n-1} 的取值, 确定下一个被测对象的特征值 id_n ;

[0017] 如果 $F_{n-1} = 1$, 则 $id_n = id_{n-1} + t_{n-1}$;

[0018] 如果 $F_{n-1} = 0$, 则 $id_n = id_{n-1} - t_{n-1}$;

[0019] 其中, $t_{n-1} = \sum_{n=2}^{L+1} [|f(id_{n-1}, x_{n-1}) - f(id_{n-2}, x_{n-2})|] |\alpha|$, L 表示测试的总个数且 $L \geq 1$;

x_{n-1} 为根据词汇特征值取 id_{n-1} 时所对应的难度系数值; x_{n-2} 为根据词汇特征值取 id_{n-2} 时所对应的难度系数值; $f(id_{n-1}, x_{n-1})$ 表示当难度系数值为 x_{n-1} 时 id_{n-1} 所对应的特征值生成函数; $f(id_{n-2}, x_{n-2})$ 表示当难度系数值为 x_{n-2} 时 id_{n-2} 所对应的特征值生成函数; α 的符号由 id_{n-1} 对应标记 F_{n-1} 的取值决定: 如果 $F_{n-1} = 1$, 则符号为正, 否则为负;

[0020] Step7、根据受试者对已经测试过的对象的特征值及其状态值得出一个 $2 \times L$ 测试结果矩阵 R :

$$[0021] \quad R = \begin{bmatrix} id_1 & id_2 & id_3 & id_4 & \dots & id_{L-1} & id_L \\ F_1 & F_2 & F_3 & F_4 & \dots & F_{L-1} & F_L \end{bmatrix}$$

[0022] 其中, L 表示测试的总个数且 $L \geq 1$;

[0023] Step8、将测试结果矩阵 R 进行处理得出受试者是否掌握测试结果:

[0024] 如果特征值 id_L 对应的 F_L 为 1 时, 则认为在 id_n 左右邻域中的词汇受试者都能掌握;

[0025] 如果特征值 id_L 对应的 F_L 为 0 时, 则认为在 id_n 左右邻域中的词汇受试者没有掌握;

[0026] Step9、根据测试结果, 得出受试者的词汇测评域估计值为:

[0027]

$$result = \sum_{k=1}^N 2 \times \psi_k(id_k) - \sum_{j=1}^M 2 \times \varphi_j(id_j)$$

[0028] 其中, N 表示测试判断对的总个数, M 表示测试判断错的总个数; $\psi_k(id_k)$ 表示被测对象特征值 id_k 对应的词汇的选项正确时的左右邻域的取值, $\varphi_j(id_j)$ 表示被测对象特征值 id_j 对应的词汇的选项错误时的左右邻域的取值。

[0029] 所述步骤 Step6 能替换为:

[0030] 如果 $F_{n-1} = 1$, 则 $id_n = id_{n-1} + t_{n-1}$;

[0031] 如果 $F_{n-1} = 0$, 则 $id_n = id_{n-1} - t_{n-1}$;

[0032] 其中, $t_{n-1} = \sum_{n=2}^{L+1} \zeta_{n-1} [|f(id_{n-1}, x_{n-1}) - f(id_{n-2}, x_{n-2})|] |\alpha|$, L 表示测试的总个数且

$L \geq 1$; ; ζ_{n-1} 表示 $[|f(id_{n-1}, x_{n-1}) - f(id_{n-2}, x_{n-2})|]$ 的加权值, $\zeta_{n-1} > \zeta_{n-2} > \zeta_{n-3} > \dots > \zeta_2 > \zeta_1$ 且 $\zeta_{n-1} + \zeta_{n-2} + \dots + \zeta_2 + \zeta_1 = 1$; 其中, α 的符号由 id_{n-1} 对应标记 F_{n-1} 的取值决定: 如果 $F_{n-1} = 1$, 则符号为正, 否则为负。

[0033] 所述特征值生成函数 $f(id_{n-2}, x_{n-2}) = \frac{id_{n-2}}{1 + Ae^{-x_{n-2}}}$; 其中, A 为函数特征系数, $n \geq 2$ 。

[0034] 本发明的工作原理是:

[0035] 第一步, 建立数学模型。

[0036] 第二步, 根据建立的数学模型: 首先确定所有词汇的属性值及进行难度级别划分; 接着随机从第 1 个难度级别的词汇中选取一个属性值作为初始特征值并确定词汇难度系数; 然后计算第一个测试词汇的对应的特征值并对其所对应的选项进行判断; 再接着计算特征值所对应该的难度系数值, 同时根据状态值确定下一个特征值; 最后根据测试评价对象矩阵处理得出受试者是否掌握测试结果并得出受试者的词汇测评域估计值。

[0037] 本发明的有益效果是:

[0038] 1、测试词汇之间存在关联性, 下一个测试词汇和已经测试了的测试词的特征值相关联, 具有 n 阶马尔可夫性质。特征值的增减量由一个与特征值相关的函数来计算, 这样使得测试词汇之间有一个关联性, 而不是以一种随机的方式来测试, 关联的程度还可以进行加权计算, 即从该测试词汇起直到以前 n 个词的测试状态的影响, 离该词远的其测试状态对现在的影响就越小。这样使得测试词汇之间的关联程度更加的紧密。大大的提高了下一个测试词汇和已经测试的结果密切的关联, 很好的克服目前测试词汇中测试词没有关联性的缺陷, 也更准确的反应被测者的真实水平词汇水平。

[0039] 2、词汇的难度系数会根据测试的判断对错进行调整, 测试对了难度系数会增加, 测试错误了难度系数会减小。难度系数值的变化也具有关联性质, 可以实时的调整难度。

[0040] 3、要测试出受试者的大概水平, 所需要测试的词汇增多, 而且测试词汇之间的随机性减小, 关联性大大增强。当测试的词汇个数少的时候, 所得出结果和实际水平相比偏低, 随着测试词汇个数的增加, 测试结果会越来越接近实际水平。

附图说明

[0041] 图 1 是本发明测试评价矩阵计算方法流程图;

[0042] 图 2 是本发明测试评价矩阵的信号模式处理图;

[0043] 图 3 是本发明测评计算方法的结果计算流程图。

具体实施方式

[0044] 实施例 1: 如图 1-3 所示, 一种基于马尔科夫的词汇量测试与评估方法, 所述方法的具体步骤如下:

[0045] 首先建立数学模型:

[0046] 已知: 对象特征值对应的状态值 $F_n = 1$ 表示测试词汇答对, $F_n = 0$ 表示测试词汇答错; 特征差值计算函数选取为 $f(B, x) = \frac{B}{1 + Ae^{a+bx}}$ 形式的函数, 实例该函数的依据是: 如

果是以英语词汇量来说, 根据国内外学者对词汇量的研究, 再结合我国和地区的英语教学, 推算出我国高级英语学习者所掌握的词汇量大概在 10000-13000 个词汇左右。所选择的样本词汇来源依据是排除所有的人名、地名, 以及一些功能 - 语法词和内容 - 实意的同源词。例如某个词汇有单数和复数, 则只取单数作为测试词汇; 某个词汇有动词和名词形式,

则选择动词形式作测试词汇；形容词和副词时，取形容词作测试词汇；以及排除一些非词汇符号，这样筛选后大概主要词汇就只有 15000 个词汇。所以把这个数据作为词汇量的测试上限。对象特征函数的最大值也不可能超这个上限，所以函数中的 $A = 20$ ，（当 A 取值大，函数值变化慢；当 A 取值小，函数值变化就快）， $a+bx = -x$ 。则最终实例化的 $f(B, x)$ 为：

$$f(id_{n-2}, x_{n-2}) = \frac{id_{n-2}}{1 + 20e^{-x_{n-2}}}, n \geq 2; \text{ 如果把 } f(B, x) \text{ 实例为更简单的函数也可以, 本实例中用}$$

上述函数来说明。对象特征值变化函数 $f(id_{n-2}, x_{n-2})$ ；对象特征值变化量 t_{n-1} ；测试评价矩阵 R ：表示受试者测试的结果矩阵；可变邻域函数 $\psi_k(id_k), \varphi_j(id_j)$ 。

[0047] 然后进行测评的算法，包括下述步骤：

[0048] Step1、确定所有词汇的属性值为 $1 \dots i (i \in N^*)$ ，并以每 1500 个词汇划分为 1 个难度级别；其中，最后小于 1500 的词汇归于最后一个难度级别；

[0049] Step2、随机从第 1 个难度级别的词汇中选取一个属性值作为初始特征值 id_0 ，并确定词汇难度系数为 x_0 ；其中， $x_0 = 1$ 表示初始的难度系数值；

[0050] Step3、计算第一个测试词汇对应的特征值： $id_1 = id_0 + [f(id_0, x_0)]$ ；其中， $f(id_0, x_0)$ 为初始特征值生成函数；

[0051] Step4、对被测对象的特征值 id_1 对应的词汇的选项进行判断：

[0052] 如果被测对象的特征值 id_1 对应的词汇的选项正确，则定义标记 $F_1 = 1$ ；

[0053] 如果被测对象的特征值 id_1 对应的词汇的选项错误，则定义标记 $F_1 = 0$ ；

[0054] Step5、计算 id_{n-1} 所对应的词汇难度系数 x_{n-1} 的取值：

[0055] 如果 $F_{n-1} = 1$ ，则 $x_{n-1} = x_{n-2} + 1$ ；

[0056] 如果 $F_{n-1} = 0$ ，则 $x_{n-1} = x_{n-2} - 0.5$ ；

[0057] 其中， $n \geq 2$ ；

[0058] Step6、根据 F_{n-1} 的取值，确定下一个被测对象的特征值 id_n ：

[0059] 如果 $F_{n-1} = 1$ ，则 $id_n = id_{n-1} + t_{n-1}$ ；

[0060] 如果 $F_{n-1} = 0$ ，则 $id_n = id_{n-1} - t_{n-1}$ ；

[0061] 其中， $t_{n-1} = \sum_{n=2}^{L+1} [|f(id_{n-1}, x_{n-1}) - f(id_{n-2}, x_{n-2})|] \alpha$ ， L 表示测试的总个数且 $L \geq 1$ ；

x_{n-1} 为根据词汇特征值取 id_{n-1} 时所对应的难度系数值； x_{n-2} 为根据词汇特征值取 id_{n-2} 时所对应的难度系数值； $f(id_{n-1}, x_{n-1})$ 表示当难度系数值为 x_{n-1} 时 id_{n-1} 所对应的特征值生成函数； $f(id_{n-2}, x_{n-2})$ 表示当难度系数值为 x_{n-2} 时 id_{n-2} 所对应的特征值生成函数； α 的符号由 id_{n-1} 对应标记 F_{n-1} 的取值决定：如果 $F_{n-1} = 1$ ，则符号为正，否则为负；

[0062] Step7、根据受试者对已经测试过的对象的特征值及其状态值得出一个 $2 \times L$ 测试结果矩阵 R ：

$$[0063] \quad R = \begin{bmatrix} id_1 & id_2 & id_3 & id_4 & \dots & id_{L-1} & id_L \\ F_1 & F_2 & F_3 & F_4 & \dots & F_{L-1} & F_L \end{bmatrix}$$

[0064] 其中， L 表示测试的总个数且 $L \geq 1$ ；

[0065] Step8、将测试结果矩阵 R 进行处理得出受试者是否掌握测试结果：

[0066] 如果特征值 id_L 对应的 F_L 为 1 时,则认为在 id_n 左右邻域中的词汇受试者都能掌握;

[0067] 如果特征值 id_L 对应的 F_L 为 0 时,则认为在 id_n 左右邻域中的词汇受试者没有掌握;

[0068] Step9、根据测试结果,得出受试者的词汇测评域估计值为:

[0069]

$$result = \sum_{k=1}^N 2 \times \psi_k(id_k) - \sum_{j=1}^M 2 \times \varphi_j(id_j)$$

[0070] 其中,N表示测试判断对的总个数,M表示测试判断错的总个数; $\psi_k(id_k)$ 表示被测对象特征值 id_k 对应的词汇的选项正确时的左右邻域的取值, $\varphi_j(id_j)$ 表示被测对象特征值 id_j 对应的词汇的选项错误时的左右邻域的取值。

[0071] 其中,所述特征值生成函数 $f(id_{n-2}, x_{n-2}) = \frac{id_{n-2}}{1 + 20e^{-x_{n-2}}}$, $n \geq 2$ 。

[0072] 具体分析如下:

[0073] 假设:测试对象的特征值就是单词的使用频率排序值,结合图 1、图 2 和图 3 的实例,受试者是测试了 8 个词汇,第 5 个错误、第 7 个错误的情况下,初始值 $id_0 = 1200$;

[0074] 如图 3 所示:测试样本难度系数值的初始值为 $x_0 = 1$,根据 $id_1 = id_0 + \lfloor f(id_0, x_0) \rfloor$ 得出第一个测试词汇所对应的特征为: $id_1 = 1200 + 143 = 1343$,并且判断 $F_1 = 1$,则 $x_1 = x_0 + 1 = 2$;

[0075] 根据步骤 Step6 计算 $id_2 = id_1 + t_1$;

[0076]

$$\begin{aligned} t_1 &= \lfloor f(id_1, x_1) - f(id_0, x_0) \rfloor \\ &= \lfloor 362 - 143 \rfloor \\ &= 219 \end{aligned}$$

[0077] 再计算 id_3 , $id_3 = id_2 + t_2$;

[0078]

$$\begin{aligned} t_2 &= \lfloor f(id_2, x_2) - f(id_1, x_1) \rfloor + t_1 \\ &= \lfloor 782 - 362 \rfloor + t_1 \\ &= 420 + 219 \\ &= 639 \end{aligned}$$

[0079] 本实例中测试的 8 个词汇的难度系数汇总后为: $x_1 = [1 \ 2 \ 3 \ 4 \ 5 \ 4.5 \ 5.5 \ 4.5 \ 5.5]$, $1 = 0, \dots, 7$;

[0080] 如图 2 所示某个被测对象的特征值对应的 $F = 1$ 则左右虚线内的单词可以,认为是掌握的,根据常用 15000 词分成 10 级,假设各级的邻域函数值 = $[500 \ 400 \ 300 \ 200 \ 100 \ 80 \ 60 \ 40 \ 20 \ 10]$;

[0081] 对象特征值生成函数 $f(id_{n-2}, x_{n-2}) = \frac{id_{n-2}}{1+20e^{-x_{n-2}}}$, $n \geq 2$, 根据图 1 则得出:

[0082] $t_m = [219 \ 639 \ 1467 \ 3089 \ 2140 \ 3402 \ 4623]$, $m = 1, \dots, 7$;

[0083] 得出 $id_L = [1343 \ 1562 \ 2201 \ 3668 \ 6757 \ 4615 \ 8017 \ 3393]$, $L = 1, \dots, 8$;

[0084] 最后得出 R 矩阵:

[0085] $R = \begin{bmatrix} 1343 & 1562 & 2201 & 3668 & 6757 & 4615 & 8017 & 3393 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 \end{bmatrix}$;

[0086] 根据图 3 中的流程一次一次进行判断, 得出算法结果:

[0087]

$$\begin{aligned} result &= \sum_{k=1}^a 2 \times \psi_k(id_k) - \sum_{j=1}^b 2 \times \varphi_j(id_j) \\ &= 2 \times 500 + 2 \times 400 + 2 \times 400 + 2 \times 300 - 2 \times 100 + 2 \times 400 - 2 \times 80 + 2 \times 300 \\ &= 3840 \end{aligned}$$

[0088] 其中, $N = 6$, $M = 2$; 被测对象特征值 id_k 对应的词汇的选项正确时的左右邻域的取值 $\psi_k(id_k) = [500, 400, 400, 300, 400, 300]$ (如: $id_1 = 1343$ 对应的词汇属于第 1 级, 第 1 级对应的邻域函数值为 500; 其它左右邻域的取值理由与其相同); 测对象特征值 id_j 对应的词汇的选项错误时的左右邻域的取值 $\varphi_j(id_j) = [100, 80]$ 。

[0089] 综上, 得出的算法结果就是该受试者最终的评价, 把该结果作为该测试者的词汇水平的估计值, 即估计出该受试者大概的词汇量为 3840 个。

[0090] 实施例 2: 如图 1-3 所示, 一种基于马尔科夫的词汇量测试与评估方法, 所述方法的具体步骤如下:

[0091] 首先建立数学模型:

[0092] 已知: 对象特征值对应的状态值 $F_n = 1$ 表示测试词汇答对, $F_n = 0$ 表示测试词汇答错; 加权值为 ζ_n ; 特征差值计算函数选取为 $f(B, x) = \frac{B}{1+Ae^{a+bx}}$ 形式的函数, 实例该

函数的依据是: 如果是以英语词汇量来说, 根据国内外学者对词汇量的研究, 再结合我国和地区的英语教学, 推算出我国高级英语学习者所掌握的词汇量大概在 10000-13000 个词汇左右。所选择的样本词汇来源依据是排除所有的人名、地名, 以及一些功能-语法词和内容-实意的同源词。例如某个词汇有单数和复数, 则只取单数作为测试词汇; 某个词汇有动词和名词形式, 则选择动词形式作测试词汇; 形容词和副词时, 取形容词作测试词汇; 以及排除一些非词汇符号, 这样筛选后大概主要词汇就只有 15000 个词汇。所以把这个数据作为词汇量的测试上限。对象特征函数的最大值也不可能超这个上限, 所以函数中的 A

$= 20$, $a+bx = -x$ 。则最终实例化的 $f(B, x)$ 为: $f(id_{n-2}, x_{n-2}) = \frac{id_{n-2}}{1+20e^{-x_{n-2}}}$, $n \geq 2$, 如果把

$f(B, x)$ 实例为更简单的函数也可以, 本实例中用上述函数来说明。对象特征值变化函数 $f(id_{n-2}, x_{n-2})$; 对象特征值变化量 t_{n-1} ; 测试评价矩阵 R: 表示受试者测试的结果矩阵; 可变邻域函数 $\psi_k(id_k)$, $\varphi_j(id_j)$ 。

[0093] 然后进行测评的算法,包括下述步骤:

[0094] Step1、确定所有词汇的属性值为 $1 \dots i (i \in \mathbb{N}^*)$,并以每 1500 个词汇划分为 1 个难度级别;其中,最后小于 1500 的词汇归于最后一个难度级别;

[0095] Step2、随机从第 1 个难度级别的词汇中选取一个属性值作为初始特征值 id_0 ,并确定词汇难度系数为 x_0 ;其中, $x_0 = 1$ 表示初始的难度系数值;

[0096] Step3、计算第一个测试词汇对应的特征值: $id_1 = id_0 + [f(id_0, x_0)]$; 其中, $f(id_0, x_0)$ 为初始特征值生成函数;

[0097] Step4、对被测对象的特征值 id_1 对应的词汇的选项进行判断:

[0098] 如果被测对象的特征值 id_1 对应的词汇的选项正确,则定义标记 $F_1 = 1$;

[0099] 如果被测对象的特征值 id_1 对应的词汇的选项错误,则定义标记 $F_1 = 0$;

[0100] Step5、计算 id_{n-1} 所对应的词汇难度系数 x_{n-1} 的取值:

[0101] 如果 $F_{n-1} = 1$,则 $x_{n-1} = x_{n-2} + 1$;

[0102] 如果 $F_{n-1} = 0$,则 $x_{n-1} = x_{n-2} - 0.5$;

[0103] 其中, $n \geq 2$;

[0104] Step6、根据 F_{n-1} 的取值,确定下一个被测对象的特征值 id_n :

[0105] 如果 $F_{n-1} = 1$,则 $id_n = id_{n-1} + t_{n-1}$;

[0106] 如果 $F_{n-1} = 0$,则 $id_n = id_{n-1} - t_{n-1}$;

[0107] 其中, $t_{n-1} = \sum_{n=2}^{L+1} \zeta_{n-1} [|f(id_{n-1}, x_{n-1}) - f(id_{n-2}, x_{n-2})|] \alpha$, L 表示测试的总个数且

$L \geq 1$; ; ζ_{n-1} 表示 $[|f(id_{n-1}, x_{n-1}) - f(id_{n-2}, x_{n-2})|]$ 的加权值, $\zeta_{n-1} > \zeta_{n-2} > \zeta_{n-3} > \dots > \zeta_2 > \zeta_1$ 且 $\zeta_{n-1} + \zeta_{n-2} + \dots + \zeta_2 + \zeta_1 = 1$; 其中, α 的符号由 id_{n-1} 对应标记 F_{n-1} 的取值决定: 如果 $F_{n-1} = 1$,则符号为正,否则为负;

[0108] Step7、根据受试者对已经测试过的对象的特征值及其状态值得出一个 $2 \times L$ 测试结果矩阵 R :

$$[0109] \quad R = \begin{bmatrix} id_1 & id_2 & id_3 & id_4 & \dots & id_{L-1} & id_L \\ F_1 & F_2 & F_3 & F_4 & \dots & F_{L-1} & F_L \end{bmatrix}$$

[0110] 其中, L 表示测试的总个数且 $L \geq 1$;

[0111] Step8、将测试结果矩阵 R 进行处理得出受试者是否掌握测试结果:

[0112] 如果特征值 id_L 对应的 F_L 为 1 时,则认为在 id_n 左右邻域中的词汇受试者都能掌握;

[0113] 如果特征值 id_L 对应的 F_L 为 0 时,则认为在 id_n 左右邻域中的词汇受试者没有掌握;

[0114] Step9、根据测试结果,得出受试者的词汇测评域估计值为:

[0115]

$$result = \sum_{k=1}^N 2 \times \psi_k(id_k) - \sum_{j=1}^M 2 \times \varphi_j(id_j)$$

[0116] 其中, N 表示测试判断对的总个数, M 表示测试判断错的总个数; $\psi_k(id_k)$ 表示被测

对象特征值 id_k 对应的词汇的选项正确时的左右邻域的取值, $\varphi_j(id_j)$ 表示被测对象特征值 id_j 对应的词汇的选项错误时的左右邻域的取值。

[0117] 其中, 所述特征值生成函数 $f(id_{n-2}, x_{n-2}) = \frac{id_{n-2}}{1+20e^{-x_{n-2}}}$, $n \geq 2$ 。

[0118] 具体分析如下:

[0119] 假设: 测试对象的特征值就是单词的使用频率排序值, 结合图 1、图 2 和图 3 的实例, 受试者是测试了 8 个词汇, 第 5 个错误、第 7 个错误的情况下, 初始值 $id_0 = 1200$;

[0120] 如图 3 所示: 测试样本难度系数值的初始值为 $x_0 = 1$, 根据 $id_1 = id_0 + [f(id_0, x_0)]$ 得出第一个测试词汇所对应的特征为: $id_1 = 1200 + 143 = 1343$, 并且判断 $F_1 = 1$, 则 $x_1 = x_0 + 1 = 2$;

[0121] 根据步骤 Step6 计算 $id_2 = id_1 + t_1$;

[0122]

$$\begin{aligned} t_1 &= [|f(id_1, x_1) - f(id_0, x_0)|] \times 1 \\ &= [362 - 143] \\ &= 219 \end{aligned}$$

[0123] 再计算 id_3 , $id_3 = id_2 + t_2$;

[0124]

$$\begin{aligned} t_2 &= [|f(id_2, x_2) - f(id_1, x_1)|] \times 0.5 + t_1 \times 0.5 \\ &= [782 - 362] \times 0.5 + t_1 \times 0.5 \\ &= 420 \times 0.5 + 219 \times 0.5 \\ &= 319 \end{aligned}$$

[0125] 本实例中测试的 8 个词汇的难度系数汇总后为: $x_1 = [1 \ 2 \ 3 \ 4 \ 5 \ 4.5 \ 5.5 \ 4.5 \ 5.5]$, $l = 0, \dots, 7$;

[0126] 如图 2 所示某个被测对象的特征值对应的 $F = 1$ 则左右虚线内的单词可以, 认为是掌握的, 根据常用 15000 词分成 10 级, 假设各级的邻域函数值 = [500 400 300 200 100 80 60 40 20 10];

[0127] 对象特征值生成函数 $f(id_{n-2}, x_{n-2}) = \frac{id_{n-2}}{1+20e^{-x_{n-2}}}$, $n \geq 2$, 根据图 1 则得出:

[0128] $t_m = [219 \ 319 \ 411 \ 320 \ 232 \ 314 \ 288]$, $m = 1, \dots, 7$;

[0129] 得出 $id_l = [1343 \ 1562 \ 1881 \ 2292 \ 2612 \ 2844 \ 3158 \ 3446]$, $L = 1, \dots, 8$;

[0130] 最后得出 R 矩阵:

[0131]
$$R = \begin{bmatrix} 1343 & 1562 & 1881 & 2292 & 2612 & 2844 & 3158 & 3446 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 \end{bmatrix};$$

[0132] 根据图 3 中的流程一次一次进行判断, 得出算法结果:

[0133]

$$\begin{aligned}
 result &= \sum_{k=1}^a 2 \times \psi_k(id_k) - \sum_{j=1}^b 2 \times \varphi_j(id_j) \\
 &= 2 \times 500 + 2 \times 400 + 2 \times 400 + 2 \times 400 - 2 \times 400 + 2 \times 400 - 2 \times 300 + 2 \times 300 \\
 &= 3400
 \end{aligned}$$

[0134] 其中, $N = 6, M = 2$; 被测对象特征值 id_k 对应的词汇的选项正确时的左右邻域的取值 $\psi_k(id_k) = [500, 400, 400, 400, 400, 300]$ (如: $id_1 = 1343$ 对应的词汇属于第 1 级, 第 1 级对应的邻域函数值为 500; 其它左右邻域的取值理由与其相同); 测对象特征值 id_j 对应的词汇的选项错误时的左右邻域的取值 $\varphi_j(id_j) = [400, 300]$ 。

[0135] 综上, 得出的算法结果就是该受试者最终的评价, 把该结果作为该测试者的词汇水平的估计值, 即估计出该受试者大概的词汇量为 3400 个。

[0136] 上面结合附图对本发明的具体实施方式作了详细说明, 但是本发明并不限于上述实施方式, 在本领域普通技术人员所具备的知识范围内, 还可以在不脱离本发明宗旨的前提下作出各种变化。

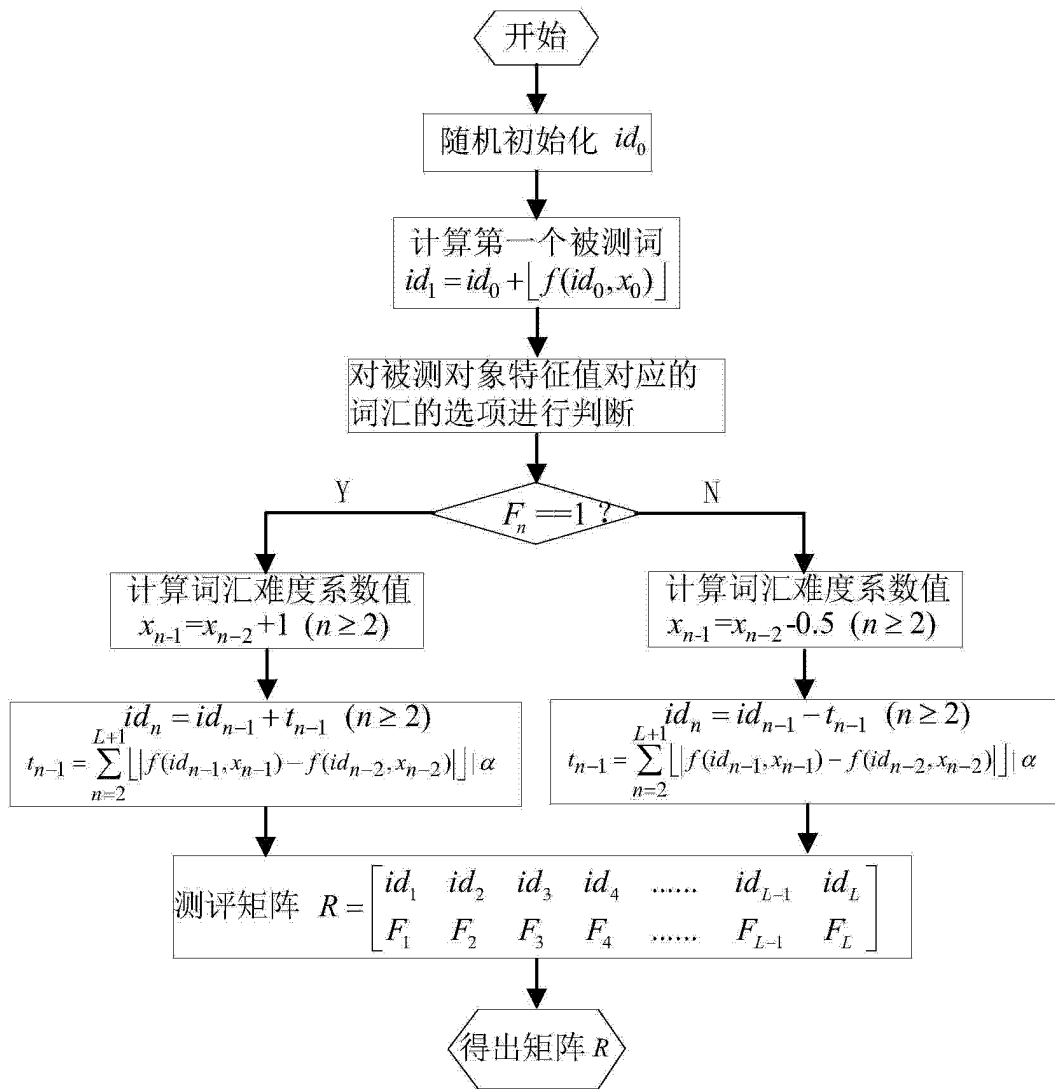


图 1

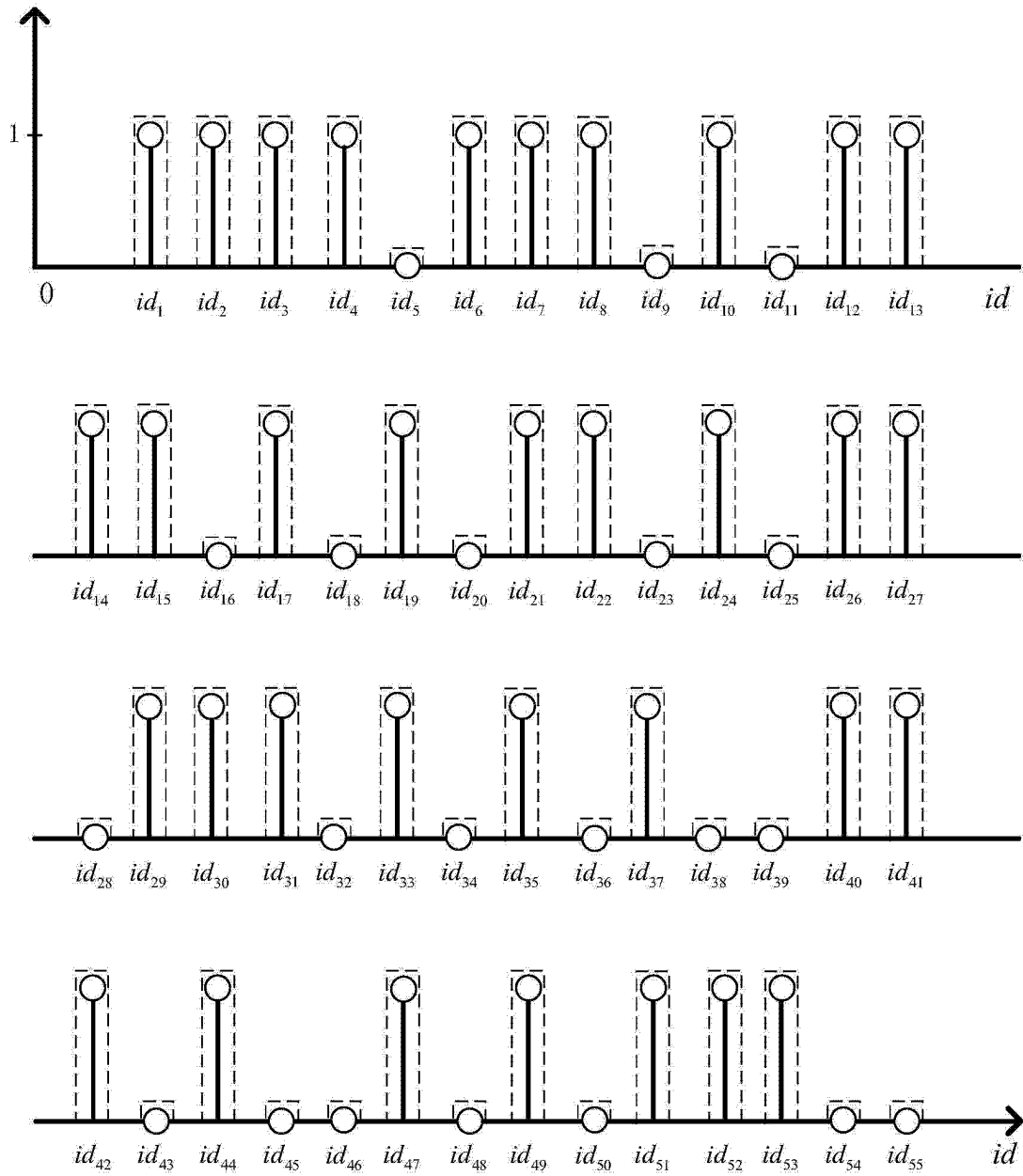


图 2

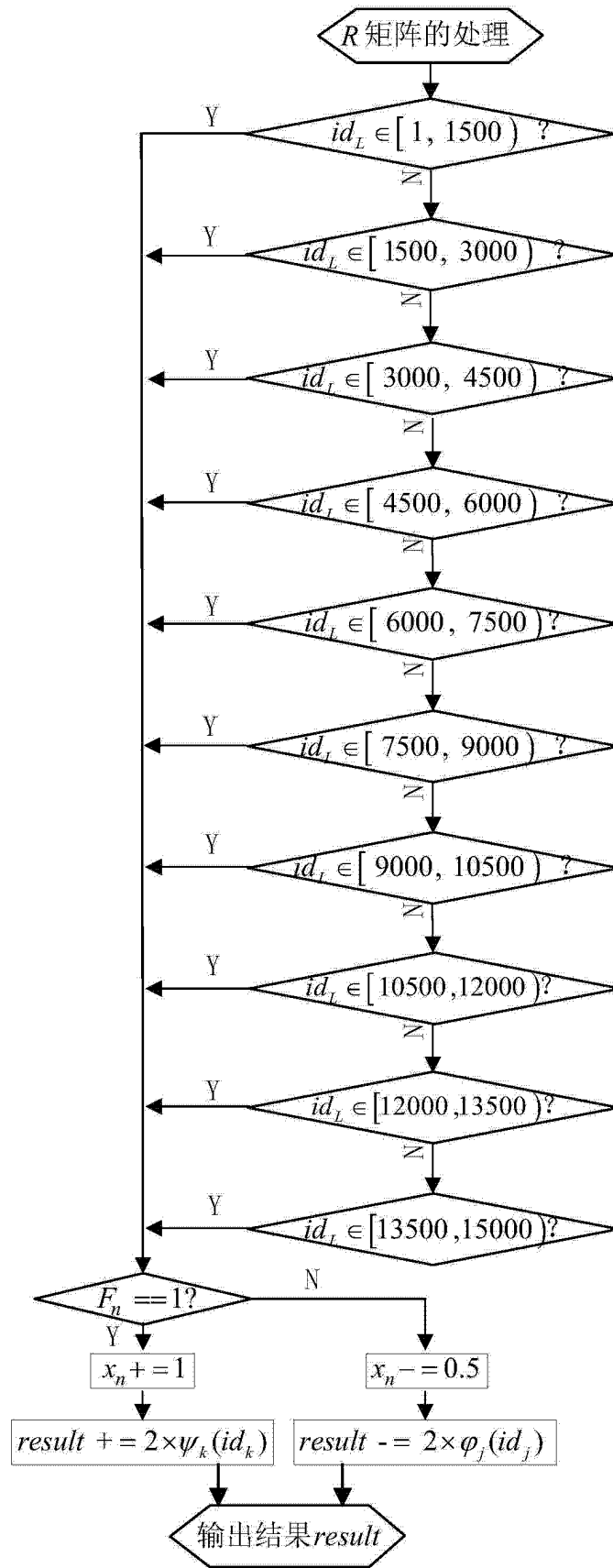


图 3