

doi: 10.3969/j.issn.1674-232X.2013.03.018

基于云模型的外语词汇量评估

陶利民

(杭州师范大学杭州国际服务工程学院, 浙江 杭州 310012)

摘要: 提出一种基于云模型的词汇量定性评估方法, 利用云模型的3个特征参数期望 Ex , 熵 En , 超熵 He 分别反映被测人员的词汇量水平、被测人员的稳定性和心理素质. 实例分析证明此方法实用性强, 评价结果直观、全面、深刻, 较传统测试方法更符合实际情况.

关键词: 云模型; 词汇量; 评估; 词汇量云; 云发生器

中图分类号: TP301

文献标志码: A

文章编号: 1674-232X(2013)03-0284-05

外语学习中的一个重要部分就是词汇学习. 词汇量的大小直接影响到学习者语言输入和输出的质量, 也影响其语用和语言交际的能力^[1]. 同时, 词汇量对于教学目标的确定、教材编写、教学组织和考试要求具有宏观指导意义, 也是教学诊断和教学评估的重要指标. 因此, 词汇教学成为了外语教学中的一个重要方面, 若这一环节不能很好地完成, 也就很难完成教学目标. 对外语学习者所掌握词汇进行测试评估, 是衡量词汇教学效果的一种重要手段. 传统的词汇量测试方法主要有词频法和词典法, 就是在词汇表中随机抽取若干词汇作为测试样本. 其中, 词频法是以词汇频率为依据进行词汇抽样, 词典法则是以一本词典为调查对象等距离抽样. 抽取好测试样本后, 采用多项选择法、翻译法、是非判断法、填空法或释义法评测外语学习者对样本词汇掌握程度; 再根据样本词汇测试结果按比例计算总体词汇量, 最终给出被测人员词汇量大小的估计值, 可用式(1)^[2]表示:

$$\text{词汇量估计值} = \frac{\text{样本词汇掌握数量} \times \text{总词汇量}}{\text{样本词汇量}} \quad (1)$$

由于测试结果会受到测试人员的稳定性、心理素质等多种因素的影响, 传统的测试方法显得粗糙简单, 不够深刻全面, 本文将采用基于云模型的定性评价方法来评估外语学习者的词汇量水平^[3].

李德毅院士在传统模糊集合理论及概率论基础上提出了云模型^[4], 借助模糊数学中隶属度函数的思想, 建立了一种定性定量不确定性转换模型. 云是用语言值表示的某个定性概念与其数值表示之间的不确定性转换模型, 用来反映自然语言中概念的不确定性. 云模型能统一表达定性概念的模糊性及随机性, 并将二者有机集成到一起, 构成定性和定量相互间的映射, 作为知识表示的基础, 从而实现了不确定语言值与定量数值之间的一种自然转换.

1 基本概念

定义 1(云) 设 U 是一个用精确数值表示的定量论域空间 $U = \{x\}$, T 是与 U 相联系的语言值. 若 U

中的定量值 x 是定性概念 T 的一次随机实现, x 对于 T 所表达的定性概念的隶属度 $\mu_T(x)$ 是一个具有稳定倾向的随机数, 隶属度在论域上的分布称之为隶属云, 简称云(Cloud). 每一个 x 称为一个云滴^[5].

$\mu_T(x)$ 在 $[0,1]$ 区间上取值, 云是从论域 U 到区间 $[0,1]$ 的映射, 即:

$$\mu_T(x):U \rightarrow [0,1], \forall x \in Ux \rightarrow \mu_T(x).$$

在进行词汇量测试评估的背景下, U 表示测试结果, 如 2 000; T 表示词汇量掌握程度, 如初级; μ 表示某一测试结果对于某一程度的隶属度.

定义 2(词汇量云) 词汇量云(Vocabulary Cloud) 是一种特殊形式的云模型, 它将测试人员所掌握词汇量用云模型的方式反映出来. 可将词汇量云定性表示成: $VC = Cloud(Ex, En, He)$. 其中, Ex, En, He 是反映词汇量云的 3 个数字特征^[6]: Ex 是词汇量期望, 是最能代表词汇量概念的点, 表明测试人员所掌握的基本词汇量. En 是词汇量熵, 反映了测试人员掌握词汇量达到某种程度所具有的不确定性. 一方面 En 反映了能够代表这个词汇量概念的云滴所具有的离散程度, 另一方面也反映了对词汇量概念是一种亦此亦彼的度量. He 是词汇量超熵, 反映了词汇量熵的不确定性, 由熵的随机性和模糊性共同决定.

若要用云模型表示“词汇量约 2 000”这一定性概念, 则可将期望 Ex 的值定为 2 000, 设 En 和 He 的值分别为 150, 20. 图 1 就是利用正向云发生器算法(算法 2)生成的语言值“词汇量约 2 000”的隶属云.

云模型具有 $3En$ 规则, 就是对于某一个定性概念或者知识, 其相应的云对象大部分位于 $[Ex - 3En, Ex + 3En]$ 之内, 位于 $[Ex - 3En, Ex + 3En]$ 之外的云滴元素只是小概率事件, 通常都可忽略. 图 1 中就体现了这一规则.

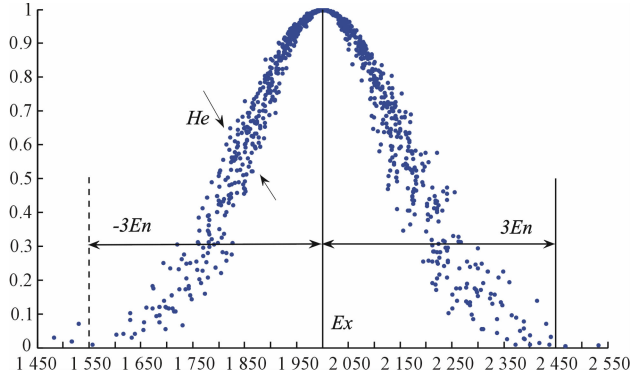


图 1 语言值“词汇量 2 000 左右”的隶属云

Fig. 1 Membership cloud of “vocabulary around 2 000”

2 评估步骤

利用云模型进行词汇量测试评估, 主要经过以下步骤:

- 1) 抽样测试. 对测试人员进行 n 次词汇抽样测试, 产生一系列的测试结果定量值: x_1, x_2, \dots, x_n .
- 2) 生成词汇量云. 将测试结果 x_i 作为 n 个云滴, 利用逆向云发生器算法^[4,7](算法 1), 生成词汇量云 VC 的 3 个数字特征 (Ex, En, He) .

算法 1 逆向云发生器算法

输入: n 个云滴 x_i .
输出: (Ex, En, He) .
算法步骤:

① 根据 x_i 计算数据的样本均值 $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$, 一阶样本中心距 $d = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{X}|$, 样本方差 $S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2$;

② $Ex = \bar{X}; En = \sqrt{\frac{\pi}{2}} \times d; He = \sqrt{S^2 - En^2}$.

3) 综合词汇量云. 若对同一测试人员作了不同级别(设有 N 个级别) 词汇的测试, 那么每一个级别会形成一个词汇量云, 共有 N 个词汇量云, $VC_i(Ex_i, En_i, He_i), i = 1, 2, \dots, N$. 此时, 需要利用式(2) 将不同级别的词汇量云综合起来得到测试人员词汇量的综合评估云(Comprehensive Assessment Cloud) $CAC(Ex, En, He)$.

$$\begin{cases} Ex = Ex_1 + Ex_2 + \cdots + Ex_i; \\ En = \sqrt{En_1^2 + En_2^2 + \cdots + En_i^2}; \quad i = 1, 2, \cdots, N; \\ He = \sqrt{He_1^2 + He_2^2 + \cdots + He_i^2}. \end{cases} \quad (2)$$

4) 结果分析. 结合综合评估云 $CAC(Ex, En, He)$, 利用正向云发生器算法^[8-9](算法 2), 对测试人员的词汇量进行定性分析.

算法 2 正向云发生器算法

输入: 表示定性概念的 3 个数字特征 (Ex, En, He) , 云滴数量 N .

输出: N 个云滴 x 及其确定度 μ , 即 $(\text{drop}(x_1, \mu_T(x_1)), \text{drop}(x_2, \mu_T(x_2)), \cdots, \text{drop}(x_N, \mu_T(x_N)))$.

算法步骤:

- ① 生成以 En 为期望值, He 为方差的一个正态随机数 $En'_i = \text{NORM}(En, He)$;
- ② 生成以 Ex 为期望值, En'_i 为方差的一个正态随机数 $x_i = \text{NORM}(Ex, En'_i)$;
- ③ 计算 $\mu_T(x_i) = e^{-\frac{(x_i - Ex)^2}{2(En'_i)^2}}$;
- ④ 令带有确定度 $\mu_T(x_i)$ 的 x_i 为数域中的一个云滴;
- ⑤ 重复 ① ~ ④, 直至产生 N 个云滴止.

当然, 也可以利用算法 2 对各个级别的测试结果作定性分析.

3 实例分析

日语能力测试(JLPT; The Japanese Language Proficiency Test)是由日本国际交流基金会及日本国际教育支援协会于 1984 年建立的一套较为完整的考试评价体系, 共分为 5 个级别(N1、N2、N3、N4、N5), 所要求的词汇量分别为 12 000、7 000、3 600、1 800、900. 本文以 N2 级所要求的词汇量作为测试样本, 将其分为 2 000 个初级词汇、2 000 个中级词汇、3 000 个高级词汇. 测试时词汇量样本的大小影响测试结果, 词汇量样本越大测试结果越准确. 考虑到测试人员的心理, 取各级别词汇的 10% 作为测试样本^[10-11]. 为了计算需要, 进行 15 组测试, 根据测试人员的测试结果, 利用式(1)可以估算出测试人员所掌握的词汇量. 表 1 列出的是 A、B 两位测试人员每组词汇的实际测试结果. 可将每组词汇测试结果看作是一个云滴, 因此每个级别的词汇测试可产生 15 个云滴.

表 1 被测人员每组词汇实际测试结果

Tab. 1 Vocabulary actual test results of tested personnel

序号	A			B		
	初级词汇	中级词汇	高级词汇	初级词汇	中级词汇	高级词汇
1	1 954	1 839	2 516	1 692	1 606	1 859
2	1 736	1 661	2 415	1 712	1 625	1 726
3	1 764	1 810	2 543	1 626	1 626	1 789
4	1 723	1 726	2 557	1 657	1 590	1 857
5	1 830	1 754	2 492	1 629	1 581	1 823
6	1 834	1 745	2 520	1 698	1 599	1 813
7	1 830	1 768	2 434	1 703	1 610	1 747
8	1 792	1 693	2 532	1 696	1 582	1 787
9	1 777	1 720	2 613	1 702	1 581	1 878
10	1 821	1 768	2 504	1 652	1 627	1 826
11	1 785	1 716	2 514	1 712	1 595	1 787
12	1 824	1 773	2 466	1 695	1 589	1 830
13	1 782	1 766	2 507	1 684	1 612	1 781
14	1 836	1 711	2 549	1 710	1 554	1 808
15	1 763	1 748	2 460	1 685	1 604	1 824

将表 1 中每个级别的测试结果当作云滴, 作为逆向云发生器算法(算法 1)的输入, 就能计算出每个被测人员的初级、中级和高级 3 个级别分级词汇量云的特征参数 (Ex, En, He) , 由此得到被测人员的各级词

汇量的定性评价结果. 结合云模型的 3 个特征参数的含义及实际应用情况,可利用这 3 个参数来反映影响被测人员测试结果的不同因素,其中,期望 Ex 最能体现被测人员的词汇量水平,熵 En 反映了被测人员的稳定性,超熵 He 反映心理素质,如表 2 所示.

表 2 分级词汇量评估结果特征参数

Tab. 2 Characteristic parameters of graded vocabulary assessment results

特征参数	A			B		
	初级	中级	高级	初级	中级	高级
Ex	1 803	1 747	2 508	1 684	1 599	1 809
En	51	42	47	28	20	39
He	21	15	19	4	3	10

根据表 2 中分级词汇量云的特征参数,利用正向云发生器算法(算法 2)可以得到被测人员 A,B 的分级词汇量评估结果的云图(图 2~图 3). 从图中可以直观地看出:被测人员 A 分级词汇量水平整体上比 B 高,但 A 测试结果较离散,稳定性不如 B. 相较而言,B 的测试结果较集中,发挥也比较稳定.

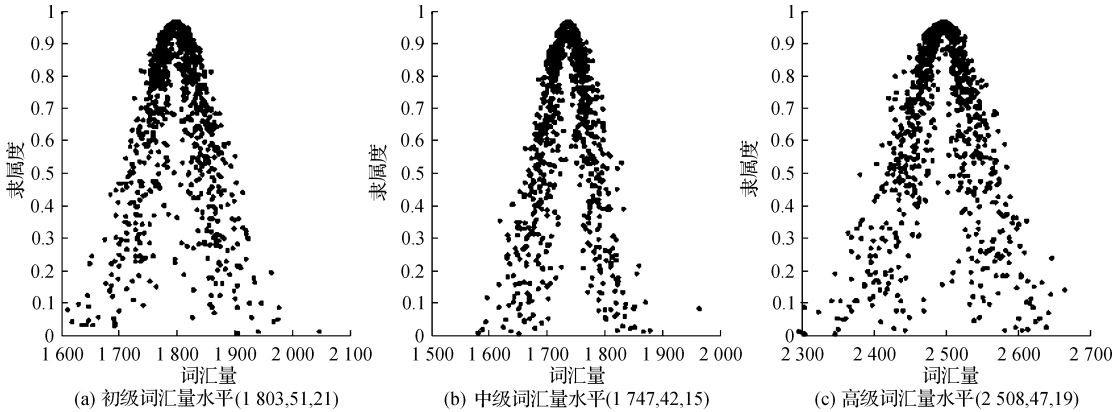


图 2 被测人员 A 分级词汇量云图

Fig. 2 Graded vocabulary cloud of tested personnel A

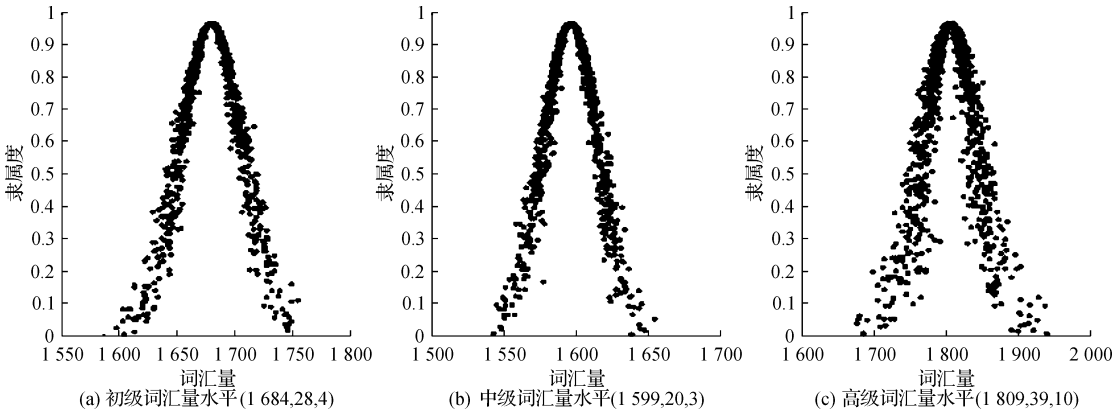


图 3 被测人员 B 分级词汇量云图

Fig. 3 Graded vocabulary cloud of tested personnel B

分级词汇量评估结果见表 2. 利用式(2)将其综合,得到被测人员的整体词汇量水平,如表 3 所示.

根据表 3 结果,利用正向云发生器算法(算法 2)可以得到被测人员的整体词汇量评估云图,如图 4 所示. 从图中可直观地看出:被测人员 A 的词汇量较大但是发挥不够稳定,比较而言,B 的词汇量中等,但是发挥较为稳定.

表 3 整体词汇量评估结果特征参数

Tab. 3 Characteristic parameters of overall vocabulary assessment results

特征参数	A	B
Ex	6 058	5 092
En	81	52
He	32	11

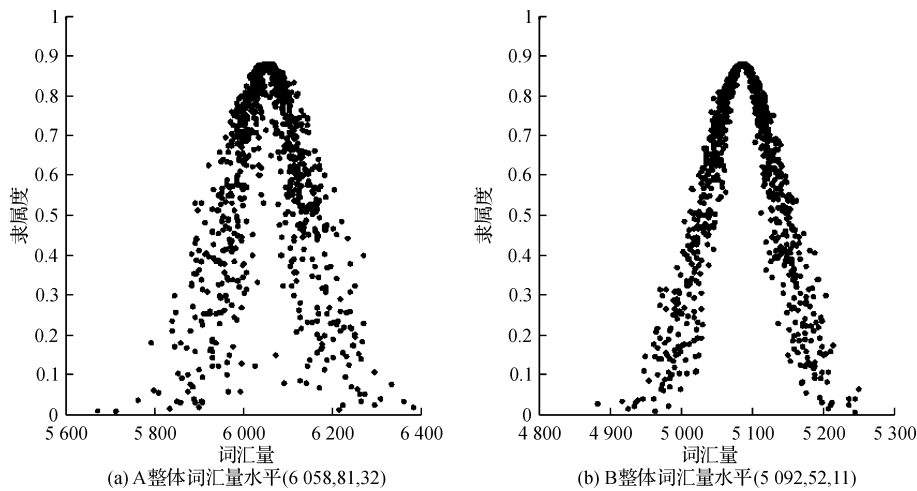


图 4 被测人员的整体词汇量水平云图
Fig. 4 Whole vocabulary level cloud of tested personnel

4 结束语

云模型能揭示自然语言中概念的两种不确定性即模糊性和随机性的内在关联性,并把两者完全集成在一起,实现定性语言值与定量数值之间的转换.与传统的定性评价方法相比较,基于云模型的定性评价方法实用性强,而且评价结果直观、全面深刻,更加符合实际情况.另外,此方法也具有可行性和有效性,用来评估外语学习者的词汇量,具有实用意义.

参考文献:

[1] 钱旭菁. 词汇量测试研究初探[J]. 世界汉语教学, 2002, 62(4): 53-62.
[2] 杨大平, 刘恒中. 英语词汇量自测与学习策略研究[J]. 山东外语教学, 2002, 91(6): 49-51.
[3] 李广琴. 词汇量对第二语言习得的影响和词汇习得策略研究[J]. 西安外国语学院学报, 2006, 14(1): 9-10.
[4] 李德毅, 刘常昱. 论正态云模型的普适性[J]. 中国工程科学, 2004, 6(8): 28-34.
[5] 杜湘瑜, 尹全军, 黄柯棣, 等. 基于云模型的定性定量转换方法及其应用[J]. 系统工程与电子技术, 2008, 30(4): 772-776.
[6] 李德毅, 孟海军, 史雪梅. 隶属云和隶属云发生器[J]. 计算机研究与发展, 1995, 32(6): 15-20.
[7] 沈进昌, 杜树新, 罗伟, 等. 基于云模型的模糊综合评价方法及应用[J]. 模糊系统与数学, 2012, 26(6): 115-123.
[8] 吕辉军, 王晔, 李德毅, 等. 逆向云在定性评价中的应用[J]. 计算机学报, 2003, 26(8): 1009-1014.
[9] 王洪利. 基于云模型的不确定性变量半定量化研究[J]. 统计与决策, 2012, 362(14): 4-6.
[10] 郁梅. 大学生英语词汇量问题研究的回顾与思考[J]. 金陵科技学院学报: 社会科学版, 2006, 20(2): 100-103.
[11] 周颂波, 顾晔, 郭亚莉. 词汇量测试述评[J]. 外语界, 2008, 124(1): 90-95.

The Assessment of Foreign Language Vocabulary Based on Cloud Model

TAO Limin

(Hangzhou Institute of Service Engineering, Hangzhou Normal University, Hangzhou 310012, China)

Abstract: The paper proposed a qualitative evaluation method of vocabulary based on cloud model. This method uses three characteristic parameters of cloud model to reflect different factors, reflect the vocabulary level of measured personnel with expectations, reflect the stability of measured personnel with entropy, reflect the psychological quality with high entropy. The analysis proves that this method is practical, and the results of evaluation are intuitive, comprehensive, and profound, more in line with actual situation than traditional methods.

Key words: cloud model; vocabulary; assessment; vocabulary cloud; cloud generator