文章编号: 1001-0920 (2016) 11-1998-07 DOI: 10.13195/j.kzyjc.2015.1060

使用在线评价信息的属性权重确定及方案排序方法

习 扬, 樊治平

(东北大学 工商管理学院，沈阳 110167)

摘 要: 提出一种使用在线评价信息的属性权重确定及方案排序方法. 首先, 将在线评价信息描述为离散型概率分布函数形式, 并构建加权累积分布函数决策矩阵; 其次, 依据该矩阵, 通过定义理想累积分布向量和每个方案与其向量的距离, 构建确定属性权重的优化模型; 再次, 依据确定的权重, 通过计算每个方案的排序值确定方案的排序结果; 最后, 依据汽车之家网站提供的汽车产品在线评价信息, 通过一个实例验证了所提出方法的实用性和可行性.

关键词: 在线评价信息；离散随机分布；属性权重；优化模型；方案排序

中图分类号: C934 **文献标志码:** A

Method for determining attribute weights and ranking alternatives based on online evaluation information

XI Yang, *FAN Zhi-ping*

(School of Business Administration，Northeastern University，Shenyang 110167，China. Correspondent：XI Yang, E-mail：xiyang_neu@163.com)

Abstract: A method for determining attribute weights and ranking alternatives is proposed by using online evaluation information. Firstly, online evaluation information with respect to each attribute is described as the form of the discrete probability distribution function, and the decision matrix about weighted cumulative distribution functions is constructed. Then, based on the decision matrix, the optimization model is constructed to determine attribute weights by defining the ideal cumulative distribution vector and the distance between each alternative and the vector. Furthermore, the ranking of alternatives can be determined through calculating the ranking value of each alternative. Finally, according to the online evaluation information of automobile products from the website ‘autohome.com.cn’, the feasibility and practicability of the proposed method is illustrated through an instance analysis.

Keywords: online evaluation information；discrete random distribution；attribute weight；optimization model；alternative ranking

0 引 言

近年来, 随着互联网和电子商务的迅猛发展, 公众或消费者给出的在线评价信息在各类网站随处可见, 例如 IT168 (http://www.it168.com/) 和汽车之家 (http://www.autohome.com.cn/) 等网站, 对消费者的购买决策影响显著[1-2]. 因此, 如何运用数据体量较大的在线评价信息进行有针对性的决策分析 (如依据大量消费者给出的在线评价信息对服务质量或产品进行排序[3-5], 对广大消费者或商品购买者提供决策支持) 是非常重要的. 需要指出的是, 由于针对方案 (如产品) 的在线评价信息通常涉及多个属性, 且在线评价信息往往具有数据体量较大的特征, 方案针对属性的在线评价信息会呈现离散随机分布的形式. 因此, 针对多属性在线评价信息具有数据体量大且呈现离散随机分布形式的特征, 在决策分析过程中, 如何确定属性的权重并进行方案的排序, 是一个新的值得关注的研究问题. 针对此问题展开研究, 发展使用在线评价信息的多属性决策理论与方法具有重要意义.

目前, 可以看到有关多属性决策理论与方法研究的丰硕成果, 已有的多属性决策方法大致可以分为两类: 一类是基于判断矩阵 (或比较矩阵) 的决策分析方法, 如 AHP 方法[6]等; 另一类是基于决策矩阵的决策分析方法, 如 TOPSIS 方法[7]等. 在多属性决策方法中, 通常需要事先确定属性的权重, 目前主要的权

收稿日期: 2015-08-21；**修回日期:** 2015-11-19.

基金项目: 国家自然科学基金项目(71571039)；中央高校基本科研业务费专项基金项目(N140607001).

作者简介: 习扬 (1990−), 男, 博士生, 从事决策理论与方法的研究；樊治平 (1961−), 男, 教授, 博士生导师, 从事运作管理、决策分析等研究.

重确定方法可归纳为3类[8]: 主观赋权法、客观赋权法和组合赋权法. 而有关属性权重的确定方法是依据所使用的决策信息类型. 通过已有研究成果可以看出, 属性权重确定方法所使用的信息类型主要包括数值[9]、区间数[10]、模糊数[11]、语言评价短语[12]等信息形式, 且权重确定方法主要有: AHP方法[6]、熵方法[13]、理想点方法[14]、优化方法[9,15]等. 此外, 有关使用在线评价信息的决策方法研究所见甚少, 但仍可看到一些相关研究成果[16-18]. 例如, 文献 [16] 通过挖掘手机产品在线评论信息得到产品的关键属性, 并使用 PROMETHEE 方法对手机产品进行排序; 文献 [17] 首先针对手机产品在线评论信息进行正向和负向情感分析, 然后应用 LDA 主题建模和多维尺度分析方法 (MDS) 确定手机品牌的定位, 并使用 TOPSIS 方法对手机产品进行排序; 文献 [18] 通过对洗护产品在线评论信息进行情感分析, 找到产品的缺陷. 但需要指出的是, 已有多属性决策分析方法以及属性权重确定方法所使用的决策信息类型很少涉及到具有离散随机分布特征的在线评价信息, 因此, 有必要研究使用在线评价信息的属性权重确定及方案排序方法.

1　原理与方法

本文考虑一个使用在线评价信息的属性权重确定及方案排序问题, 为便于分析, 采用以下符号来描述该问题中所涉及的集和量:

$C=\{C_1,C_2,\cdots,C_n\}$ 为属性集, C_j 表示第 j 个属性, $j=1,2,\cdots,n$.

$A=\{A_1,A_2,\cdots,A_m\}$ 为方案集, A_i 表示第 i 个方案, $i=1,2,\cdots,m$.

$H_j=\{H_j^1,H_j^2,\cdots,H_j^g\}$ 为在线评价过程中针对属性 C_j 所使用的评价标度集, H_j^s 表示针对属性 C_j 的第 s 个评价标度, $s=1,2,\cdots,g$, 通常 s 越大, H_j^s 所表示的评价等级越高. 例如, 在汽车之家网站中, 关于汽车产品 (方案) 的在线评价所使用的评价标度集是 5 分制评价等级形式, 即 $H_j=\{H_j^1=1,\ H_j^2=2,\ H_j^3=3,\ H_j^4=4,\ H_j^5=5\}$, 其中 1 分表示最差, 5 分表示最好. 这里, 依据大量的实际网站, 可以考虑 H_j^s 是分值的形式, $H_j^s>0$ 且 $H_j^s<H_j^{s+1}$, $s,s+1\in\{1,2,\cdots,g\}$.

K_{ij} 为参与方案 A_i 针对属性 C_j 的在线评价的用户数, $i=1,2,\cdots,m$, $j=1,2,\cdots,n$.

K_{ij}^s 为在线评价过程中使用评价标度 H_j^s 进行方案 A_i 针对属性 C_j 的评价的用户数, 其中 $i=1,2,\cdots,m$, $j=1,2,\cdots,n$, $s=1,2,\cdots,g$.

$Y_{k_{ij}}$ 为第 k_{ij} 个用户给出方案 A_i 针对属性 C_j 的评价值, $i=1,2,\cdots,m$, $j=1,2,\cdots,n$, $k_{ij}\in\{1,2,\cdots,K_{ij}\}$, $Y_{k_{ij}}\in H_j$.

本文要解决的问题是: 依据来自网站的已知在线评价信息, 如何确定属性的权重并进行方案的排序.

为了解决上述问题, 这里给出一个使用在线评价信息的属性权重确定及方案排序方法. 该方法的内容可分为 5 个部分, 下面给出每个部分的描述.

1) 构建加权累积分布函数决策矩阵.

通常, 参与在线评价的用户数 (如 K_{ij}) 较大, 并且依据评价标度集 H_j 的方案 A_i 针对属性 C_j 的评价值 (如 $Y_{k_{ij}}$) 往往具有多种可能的评价结果, 所以某一方案针对某一属性的评价结果往往会呈现离散概率分布的形式. 为方便起见, 记 X_{ij} 表示方案 A_i 针对属性 C_j 的评价结果, 显然, X_{ij} 是一个离散型随机变量. 若考虑 K_{ij} 个用户使用评价标度集 H_j 针对属性 C_j 对方案 A_i 进行评价, 则关于评价结果的离散概率分布函数可描述为

$$P_{ij}(x)=\begin{cases}\dfrac{K_{ij}^1}{K_{ij}},\ x=H_j^1;\\ \dfrac{K_{ij}^2}{K_{ij}},\ x=H_j^2;\\ \ \vdots\\ \dfrac{K_{ij}^s}{K_{ij}},\ x=H_j^s;\\ \ \vdots\\ \dfrac{K_{ij}^g}{K_{ij}},\ x=H_j^g;\end{cases}\tag{1}$$

$$i=1,2,\cdots,m,\ j=1,2,\cdots,n,\ s=1,2,\cdots,g.$$

其中: K_{ij}^s 表示在线评价过程中使用评价标度 H_j^s 进行方案 A_i 针对属性 C_j 的评价的用户数. 针对由式 (1) 描述的 $P_{ij}(x)$, 相应的累积分布函数可写为

$$F_{ij}(x)=\begin{cases}0,\ x<H_j^1;\\ \dfrac{K_{ij}^1}{K_{ij}},\ H_j^1\leqslant x<H_j^2;\\ \dfrac{K_{ij}^1+K_{ij}^2}{K_{ij}},\ H_j^2\leqslant x<H_j^3;\\ \ \vdots\\ \dfrac{K_{ij}^1+K_{ij}^2+\cdots+K_{ij}^{s-1}}{K_{ij}},\ H_j^{s-1}\leqslant x<H_j^s;\\ \ \vdots\\ 1,\ x\geqslant H_j^g;\end{cases}\tag{2}$$

$$i=1,2,\cdots,m,\ j=1,2,\cdots,n,\ s=1,2,\cdots,g.$$

进一步地, 可以构建累积分布函数决策矩阵, 即

$$[F_{ij}(x)]_{m\times n}=\begin{bmatrix} F_{11}(x) & F_{12}(x) & \cdots & F_{1n}(x)\\ F_{21}(x) & F_{22}(x) & \cdots & F_{2n}(x)\\ \vdots & \vdots & \ddots & \vdots\\ F_{m1}(x) & F_{m2}(x) & \cdots & F_{mn}(x)\end{bmatrix}.$$

设 $\boldsymbol{w}=(w_1,w_2,\cdots,w_n)$ 为属性权重向量, 其中 w_j 为属性 C_j 的权重或重要程度, 满足 $0\leqslant w_j\leqslant 1$, $\sum\limits_{j=1}^{n}w_j=1$, $j=1,2,\cdots,n$. 考虑 $[F_{ij}(x)]_{m\times n}$ 和 $\boldsymbol{w}$, 可以构建加权累积分布函数决策矩阵, 即

$$[w_jF_{ij}(x)]_{m\times n}=\begin{bmatrix} w_1F_{11}(x) & w_2F_{12}(x) & \cdots & w_nF_{1n}(x)\\ w_1F_{21}(x) & w_2F_{22}(x) & \cdots & w_nF_{2n}(x)\\ \vdots & \vdots & \ddots & \vdots\\ w_1F_{m1}(x) & w_2F_{m2}(x) & \cdots & w_nF_{mn}(x)\end{bmatrix}.$$

2) 定义理想累积分布向量和每个方案与其向量的距离.

依据构建的累积分布函数决策矩阵 $[F_{ij}(x)]_{m\times n}$, 给出关于理想累积分布向量的定义.

定义 1 理想累积分布向量被定义为

$$F^*=(F_1^*(x),F_2^*(x),\cdots,F_n^*(x)),$$

其中 $F_j^*(x)$ 是针对属性 C_j 的理想累积分布函数, 即

$$F_j^*(x)=\overline{\min}\,\{F_{ij}(x)|i=1,\cdots,m\}=\begin{cases}0,\ x<H_j^1;\\ \min\left\{\dfrac{K_{ij}^1}{K_{ij}}|i=1,2,\cdots,m\right\},\ H_j^1\leqslant x<H_j^2;\\ \min\left\{\dfrac{K_{ij}^1+K_{ij}^2}{K_{ij}}|i=1,2,\cdots,m\right\},\ H_j^2\leqslant x<H_j^3;\\ \vdots\\ \min\left\{\dfrac{K_{ij}^1+K_{ij}^2+\cdots+K_{ij}^{s-1}}{K_{ij}}|i=1,2,\cdots,m\right\},\\ \quad H_j^{s-1}\leqslant x<H_j^s;\\ \vdots\\ 1,\ x\geqslant H_j^g;\end{cases}$$

$$j=1,2,\cdots,n. \tag{3}$$

其中 $\overline{\min}$ 表示针对 $\forall x\in H_j$ 的最小函数值的算子.

采用一个图示来说明定义 1. 假设两个方案 A_1 和 A_2 针对属性 C_j 评价的累积分布函数为 $F_{1j}(x)$ 和 $F_{2j}(x)$, 则基于 $F_{1j}(x)$ 和 $F_{2j}(x)$ 确定的理想累积分布函数 $F_j^*(x)$ 如图 1 所示.

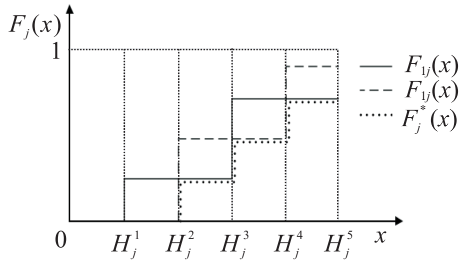

图 1 确定理想累积分布函数 $F_j^*(x)$ 的示意图

定义 2 依据加权累积分布函数决策矩阵 $[w_jF_{ij}(x)]_{m\times n}$, 加权理想累积分布向量被定义为

$$\boldsymbol{w}F^*=(w_1F_1^*(x),w_2F_2^*(x),\cdots,w_nF_n^*(x)),$$

其中 $w_jF_j^*(x)$ 是针对属性 C_j 的加权理想累积分布函数, $j=1,2,\cdots,n$.

定义 3 方案 A_i 针对属性 C_j 的评价结果 X_{ij} 的累积分布函数 $F_{ij}(x)$ 与针对属性 C_j 的理想累积分布函数 $F_j^*(x)$ 的距离被定义为

$$D(F_j^*(x),F_{ij}(x))=\sum_{s=1}^{g}(F_j^*(H_j^s)-F_{ij}(H_j^s))^2,$$

$$i=1,2,\cdots,m,\ j=1,2,\cdots,n, \tag{4}$$

其中 $F_{ij}(H_j^s)$ 和 $F_j^*(H_j^s)$ 表示累积分布函数 $F_{ij}(x)$ 和 $F_j^*(x)$ 在 $x=H_j^s$ 处的取值, $s=1,2,\cdots,g$.

定义 4 依据加权累积分布函数决策矩阵 $[w_jF_{ij}(x)]_{m\times n}$, 方案 A_i 与加权理想累积分布向量的距离被定义为

$$D_i=\sum_{j=1}^{n}D(w_jF_j^*(x),w_jF_{ij}(x))=\sum_{j=1}^{n}\sum_{s=1}^{g}(w_jF_j^*(H_j^s)-w_jF_{ij}(H_j^s))^2, \tag{5}$$

$$i=1,2,\cdots,m.$$

式 (5) 可以进一步写为

$$D_i=\sum_{j=1}^{n}r_{ij}w_j^2,\ i=1,2,\cdots,m. \tag{6}$$

其中

$$r_{ij}=\sum_{s=1}^{g}(F_j^*(H_j^s)-F_{ij}(H_j^s))^2, \tag{7}$$

$$i=1,2,\cdots,m,\ j=1,2,\cdots,n.$$

3) 构建确定属性权重的优化模型.

显然, 由式 (6) 可知, w_j 的选取应使距离 D_i 越小越好. 基于此, 可构建如下多目标最优化模型:

$$V-\min D=\{D_1,D_2,\cdots,D_m\}. \tag{8}$$

$$\text{s.t.}\ \sum_{j=1}^{n}w_j=1; \tag{9}$$

$$w_j\geqslant 0,\ j=1,2,\cdots,n. \tag{10}$$

为了便于分析和求解, 对各个方案不存在任何偏

好关系, 并且它们公平竞争, 所以可对式 (8) 所示的多个目标函数进行综合, 即将上述多目标最优化模型转化为如下单目标最优化模型:

$$\min\ Z=\sum_{i=1}^{m}D_i=\sum_{i=1}^{m}\sum_{j=1}^{n}r_{ij}w_j^2. \tag{11}$$

$$\text{s.t.}\ \sum_{j=1}^{n}w_j=1; \tag{12}$$

$$w_j\geqslant 0,\ j=1,2,\cdots,n. \tag{13}$$

4) 求解优化模型并确定属性权重.

定理 1　优化模型 (11) ～ (13) 的最优解是

$$w_j=\frac{1}{\Big(\sum\limits_{j=1}^{n}\frac{1}{\sum\limits_{i=1}^{m}r_{ij}}\Big)\Big(\sum\limits_{i=1}^{m}r_{ij}\Big)},\ j=1,2,\cdots,n. \tag{14}$$

证明　针对优化模型 (11) ～ (13), 在先不考虑非负约束条件 (13) 的情况下, 构造拉格朗日函数为

$$L(\boldsymbol{w},\lambda)=\sum_{i=1}^{m}\sum_{j=1}^{n}r_{ij}w_j^2+\lambda\Big(\sum_{j=1}^{n}w_j-1\Big), \tag{15}$$

其中 λ 为拉格朗日乘数. 对拉格朗日函数 $L(\boldsymbol{w},\lambda)$ 求关于 w_j 和 λ 的偏导数, 并使其等于 0, 即

$$\frac{\partial L(\boldsymbol{w},\lambda)}{\partial w_j}=2w_j\sum_{i=1}^{m}r_{ij}+\lambda=0,\ j=1,2,\cdots,n; \tag{16}$$

$$\frac{\partial L(\boldsymbol{w},\lambda)}{\partial \lambda}=\sum_{j=1}^{n}w_j-1=0. \tag{17}$$

联立求解式 (16) 和 (17), 若存在 $\sum_{i=1}^{m}r_{ij}\neq 0$, 则可得

$$w_j^*=\frac{1}{\Big(\sum\limits_{j=1}^{n}\frac{1}{\sum\limits_{i=1}^{m}r_{ij}}\Big)\Big(\sum\limits_{i=1}^{m}r_{ij}\Big)},\ j=1,2,\cdots,n; \tag{18}$$

$$\lambda^*=\frac{-2}{\sum\limits_{j=1}^{n}\frac{1}{\sum\limits_{i=1}^{m}r_{ij}}}. \tag{19}$$

进一步地, 对拉格朗日函数 $L(\boldsymbol{w},\lambda)$ 求关于 w_j 的 Hessian 矩阵, 即

$$H(\boldsymbol{w},\lambda)=\nabla_w^2L(\boldsymbol{w},\lambda)=$$

$$\begin{bmatrix}\frac{\partial^2L}{\partial w_1^2} & \frac{\partial^2L}{\partial w_1w_2} & \cdots & \frac{\partial^2L}{\partial w_1w_n}\\ \frac{\partial^2L}{\partial w_2w_1} & \frac{\partial^2L}{\partial w_2^2} & \cdots & \frac{\partial^2L}{\partial w_2w_n}\\ \vdots & \vdots & \ddots & \vdots\\ \frac{\partial^2L}{\partial w_nw_1} & \frac{\partial^2L}{\partial w_nw_2} & \cdots & \frac{\partial^2L}{\partial w_n^2}\end{bmatrix}=$$

$$\begin{bmatrix}2\sum\limits_{i=1}^{m}r_{i1} & 0 & \cdots & 0\\ 0 & 2\sum\limits_{i=1}^{m}r_{i2} & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & 2\sum\limits_{i=1}^{m}r_{in}\end{bmatrix}. \tag{20}$$

由式 (20) 可以看出, 若存在 $\sum_{i=1}^{m}r_{ij}\neq 0$, 则必有 $\sum_{i=1}^{m}r_{ij}>0$, 这样, 可知 $H(\boldsymbol{w},\lambda)$ 恒为正定矩阵. 因此, $L(\boldsymbol{w},\lambda)$ 有全局极小点和全局极小值. 另外, 由式 (18) 可知, 若存在 $\sum_{i=1}^{m}r_{ij}\neq 0$, 则有 $w_j^*>0$, 即满足非负约束条件 (13). 因此, 由式 (18) 确定的 w_j 是优化模型 (11) ～ (13) 的最优解. □

注 1　在现实中, $\sum_{i=1}^{m}r_{ij}=0$ 的情况几乎是不存在的. 这是因为: 如果 $\sum_{i=1}^{m}r_{ij}=0$, 则由 $r_{ij}=\sum_{s=1}^{g}(F_j^*(H_j^s)-F_{ij}(H_j^s))^2$ 可以看出, 将会出现 $F_{1j}(x)=F_{2j}(x)=\cdots=F_{mj}(x)$, 这意味着 $P_{1j}(x)=P_{2j}(x)=\cdots=P_{mj}(x)$, 即不同方案针对属性 C_j 的在线评价的离散概率分布是完全一样的, 而在现实中难以出现这种情况.

注 2　由式 (18) 确定的属性权重 w_j 与 $\sum_{i=1}^{m}r_{ij}$ 具有相关性, $\sum_{i=1}^{m}r_{ij}$ 越小, 意味着所有方案针对属性 C_j 的累积分布函数与理想累积分布函数 $F_j^*(x)$ 之间的距离之和越小, 则权重 w_j 越大. 这是因为: 在式 (18) 中, $\sum_{j=1}^{n}\frac{1}{\sum_{i=1}^{m}r_{ij}}$ 是一个常数, 与 j 没有关系, 所以 w_j 与 $\sum_{i=1}^{m}r_{ij}$ 成反比关系.

5) 确定方案排序结果.

将式 (18) 确定的属性权重代入 (6), 可得

$$D_i=\sum_{j=1}^{n}\Bigg\{r_{ij}\Bigg[\frac{1}{\Big(\sum\limits_{j=1}^{n}\frac{1}{\sum\limits_{k=1}^{m}r_{kj}}\Big)\Big(\sum\limits_{k=1}^{m}r_{kj}\Big)}\Bigg]^2\Bigg\},\quad i=1,2,\cdots,m. \tag{21}$$

显然, D_i 可以作为方案 A_i 的排序值, D_i 越小, 相应的方案 A_i 越排在前面.

综上所述, 下面给出使用在线评价信息的属性权

重确定及方案排序方法的计算步骤:

Step 1: 依据式(1)和(2), 将每个属性的在线评价信息描述为离散型概率分布函数的形式, 进而构建加权累积分布函数决策矩阵.

Step 2: 依据式(3)和定义2, 构建理想累积分布向量和加权理想累积分布向量.

Step 3: 依据式(4)～(7), 构建出确定属性权重的优化模型(11)～(13).

Step 4: 通过求解模型(11)～(13), 即依据式(18), 可得到属性权重向量 $\boldsymbol{w}^* = (w_1^*, w_2^*, \cdots, w_n^*)$.

Step 5: 依据式(21), 计算出每个方案的排序值, 通过比较所有方案的排序值可确定方案的排序结果.

2 实例分析

为了说明前文给出的使用在线评价信息的属性权重确定及方案排序方法, 这里采用汽车之家网站提供的在线评价信息对多款运动型多用途汽车(SUV)进行排序, 以便对潜在的汽车购买者提供服务与决策支持. 依据汽车之家网站, 展示汽车产品特性的属性有8个, 即: 空间(C_1)、动力(C_2)、操控(C_3)、油耗(C_4)、舒适性(C_5)、外观(C_6)、内饰(C_7)和性价比(C_8). 这里, 选择价格和产品定位相似的5款SUV汽车, 即: 吉普指南者(A_1)、马自达CX5(A_2)、斯巴鲁森林人(A_3)、丰田汉兰达(A_4)和雪佛兰科帕奇(A_5). 从汽车之家网站可以看到这5款汽车针对8个属性的大量消费者在线评价信息, 并且在线评价信息是消费者基于评价标度集 $H_j = \{H_j^1 = 1,\ H_j^2 = 2,\ H_j^3 = 3,\ H_j^4 = 4,\ H_j^5 = 5\}$ 给出的, 其中1分表示最差, 5分表示最好. 由汽车之家网站可知, 针对 $A_1 \sim A_5$, 这5款汽车进行在线评价的人数分别为: 1355、1785、1494、1425和1263. 运用Excel中的网页爬虫功能来获取相应的在线评价信息, 同时基于Excel中的Visual Basic模块编写一个宏程序, 将获取的在线评价信息结构化. 表1所示为针对5款SUV汽车在8个属性下获取的消费者在线评价信息(即依据评价标度集参与评价的用户数).

为了对上述5款SUV汽车进行排序, 采用前文给出的方法. 下面给出有关部分计算过程和结果的描述.

依据表1和式(1), 在线评价结果可以描述为离散型概率分布函数的形式, 如表2所示.

依据表2和式(2), 可得到每款汽车针对每个属性的评价结果 X_{ij} 的累积分布函数, 即 $F_{ij}(x)$, $i = 1, 2, \cdots, 5$, $j = 1, 2, \cdots, 8$. 为节省本文的篇幅, 仅以 A_1 针对 C_1 的评价结果 X_{11} 的累积分布函数 $F_{11}(x)$ 为例, 简要说明其计算过程.

表1 针对5款SUV汽车在8个属性下的参与评价的用户数

A_i	H_j	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
A_1	1	6	1	1	6	2	1	6	4
	2	21	14	3	53	25	1	58	12
	3	256	163	48	282	305	46	392	132
	4	790	652	382	572	632	208	655	573
	5	282	525	921	442	391	1 099	244	634
A_2	1	5	6	6	7	6	6	12	9
	2	5	13	2	3	13	2	30	3
	3	66	157	5	20	283	68	464	53
	4	981	908	106	198	979	378	975	480
	5	728	701	1 666	1 557	504	1 331	304	1 240
A_3	1	0	0	1	3	2	0	10	2
	2	2	5	0	15	22	3	82	23
	3	46	75	12	139	358	122	478	210
	4	385	651	286	538	714	661	724	695
	5	1 061	763	1 195	799	398	708	200	564
A_4	1	0	5	5	1	0	1	17	4
	2	0	7	20	27	4	0	98	7
	3	2	199	228	173	71	50	569	84
	4	29	770	689	564	493	424	530	521
	5	1 394	444	483	660	857	950	211	809
A_5	1	0	4	3	13	1	0	0	0
	2	1	19	4	88	8	5	13	4
	3	6	378	83	456	143	91	226	24
	4	121	720	560	582	659	581	734	233
	5	1 135	142	613	124	452	586	290	1 002

由表2可知, A_1 针对 C_1 的评价结果的离散型概率分布函数为

$$P_{11}(x) = \begin{cases} 0.004\,5, & x = 1; \\ 0.015\,4, & x = 2; \\ 0.188\,9, & x = 3; \\ 0.583\,1, & x = 4; \\ 0.208\,1, & x = 5. \end{cases}$$

依据式(2), $P_{11}(x)$ 的相应累积分布函数 $F_{11}(x)$ 可以写为

$$F_{11}(x) = \begin{cases} 0, & x < 1; \\ 0.004\,5, & 1 \leqslant x < 2; \\ 0.019\,9, & 2 \leqslant x < 3; \\ 0.208\,8, & 3 \leqslant x < 4; \\ 0.791\,9, & 4 \leqslant x < 5; \\ 1, & x \geqslant 5. \end{cases}$$

进一步地, 依据确定的 $F_{ij}(x)$, 构建出累积分布函数决策矩阵 $[F_{ij}(x)]_{5\times 8}$. 基于 $[F_{ij}(x)]_{5\times 8}$ 和定义1, 可构建理想累积分布向量为

$$F^* = (F_1^*(x), F_2^*(x), F_3^*(x), F_4^*(x), \\ F_5^*(x), F_6^*(x), F_7^*(x), F_8^*(x)).$$

其中

表 2　针对 5 款 SUV 汽车在 8 个属性下的具有离散型概率分布函数形式的评价结果

A_i	H_j	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8
	1	0.004 5	0.000 7	0.000 7	0.004 4	0.001 5	0.000 7	0.004 4	0.003
	2	0.015 4	0.010 3	0.002 2	0.039 1	0.018 5	0.000 7	0.042 8	0.008 9
A_1	3	0.188 9	0.120 3	0.035 4	0.208 1	0.225 1	0.033 9	0.289 3	0.097 4
	4	0.583 1	0.481 2	0.281 9	0.422 1	0.466 4	0.153 5	0.483 4	0.422 9
	5	0.208 1	0.387 5	0.679 7	0.326 2	0.288 6	0.811 1	0.180 1	0.467 9
	1	0.002 8	0.003 4	0.003 4	0.003 9	0.003 4	0.003 4	0.006 7	0.005
	2	0.002 8	0.007 3	0.001 1	0.001 7	0.007 3	0.001 1	0.016 8	0.001 7
A_2	3	0.037	0.088	0.002 8	0.011 2	0.158 5	0.038 1	0.259 9	0.029 7
	4	0.549 6	0.508 7	0.059 4	0.111	0.548 5	0.211 8	0.546 2	0.268 9
	5	0.407 8	0.392 7	0.933 3	0.872 3	0.282 4	0.745 7	0.170 3	0.694 7
	1	0	0	0.000 7	0.002 0	0.001 3	0	0.006 7	0.001 3
	2	0.001 3	0.003 3	0	0.01	0.014 7	0.002	0.054 9	0.015 4
A_3	3	0.030 7	0.050 2	0.008	0.093	0.239 6	0.081 7	0.319 9	0.140 6
	4	0.257 7	0.435 7	0.191 4	0.360 1	0.477 9	0.442 4	0.484 6	0.465 2
	5	0.710 2	0.510 7	0.799 9	0.534 8	0.266 4	0.473 9	0.133 9	0.377 5
	1	0	0.003 5	0.003 5	0.000 7	0	0.000 7	0.011 9	0.002 8
	2	0	0.004 9	0.014	0.018 9	0.002 8	0	0.068 8	0.004 9
A_4	3	0.001 4	0.139 6	0.16	0.121 4	0.049 8	0.035 1	0.399 3	0.058 9
	4	0.020 4	0.540 4	0.483 5	0.395 8	0.346	0.297 5	0.371 9	0.365 6
	5	0.978 2	0.311 6	0.338 9	0.463 2	0.601 4	0.666 7	0.148 1	0.567 7
	1	0	0.003 2	0.002 4	0.010 3	0.000 8	0	0	0
	2	0.000 8	0.015	0.003 2	0.069 7	0.006 3	0.004	0.010 3	0.003 2
A_5	3	0.004 8	0.299 3	0.065 7	0.361	0.113 2	0.072 1	0.178 9	0.019
	4	0.095 8	0.570 1	0.443 4	0.460 8	0.521 8	0.46	0.581 2	0.184 5
	5	0.898 7	0.112 4	0.485 4	0.098 2	0.357 9	0.464	0.229 6	0.793 3

$$F_1^*(x)=\begin{cases}0,\ x<3;\\0.001\,4,,\ 3\leqslant x<4;\\0.021\,8,\ 4\leqslant x<5;\\1,\ x\geqslant 5.\end{cases}$$

$$F_2^*(x)=\begin{cases}0,\ x<2;\\0.003\,3,\ 2\leqslant x<3;\\0.053\,5,\ 3\leqslant x<4;\\0.489\,2,\ 4\leqslant x<5;\\1,\ x\geqslant 5.\end{cases}$$

$$F_3^*(x)=\begin{cases}0,\ x<1;\\0.000\,7,\ 1\leqslant x<3;\\0.007\,3,\ 3\leqslant x<4;\\0.066\,7,\ 4\leqslant x<5;\\1,\ x\geqslant 5.\end{cases}$$

$$F_4^*(x)=\begin{cases}0,\ x<1;\\0.000\,7,\ 1\leqslant x<2;\\0.005\,6,\ 2\leqslant x<3;\\0.016\,8,\ 3\leqslant x<4;\\0.127\,8,\ 4\leqslant x<5;\\1,\ x\geqslant 5.\end{cases}$$

$$F_5^*(x)=\begin{cases}0,\ x<2;\\0.002\,8,\ 2\leqslant x<3;\\0.052\,6,\ 3\leqslant x<4;\\0.398\,6,\ 4\leqslant x<5;\\1,\ x\geqslant 5.\end{cases}$$

$$F_6^*(x)=\begin{cases}0,\ x<2;\\0.000\,7,\ 2\leqslant x<3;\\0.035\,3,\ 3\leqslant x<4;\\0.188\,8,\ 4\leqslant x<5;\\1,\ x\geqslant 5.\end{cases}$$

$$F_7^*(x)=\begin{cases}0,\ x<2;\\0.010\,3,\ 2\leqslant x<3;\\0.189\,2,\ 3\leqslant x<4;\\0.770\,4,\ 4\leqslant x<5;\\1,\ x\geqslant 5.\end{cases}$$

$$F_8^*(x)=\begin{cases}0,\ x<2;\\0.003\,2,\ 2\leqslant x<3;\\0.022\,2,\ 3\leqslant x<4;\\0.206\,7,\ 4\leqslant x<5;\\1,\ x\geqslant 5.\end{cases}$$

依据上述确定的信息并通过式(10), 可得到8个属性的权重分别为: $w_1^*=0.0469$, $w_2^*=0.1556$, $w_3^*=0.0730$, $w_4^*=0.0339$, $w_5^*=0.1046$, $w_6^*=0.1857$, $w_7^*=0.2672$, $w_8^*=0.1331$.

最后, 将所得属性权重代入式(20), 计算5款SUV汽车的排序值, 计算结果为: $D_1=0.0080$, $D_2=0.0036$, $D_3=0.0129$, $D_4=0.0119$, $D_5=0.0124$. 通过比较5 款SUV汽车的排序值可得到排序结果为 $A_2\succ A_1\succ A_4\succ A_5\succ A_3$.

3 结 论

本文给出了一种使用在线评价信息的属性权重确定及方案排序方法. 在该方法中, 将方案针对属性的在线评价信息描述为离散型概率分布函数形式, 并构建加权累积分布函数决策矩阵. 进一步地, 通过构建确定属性权重的优化模型并对模型进行求解, 可确定每个属性的权重和计算每个方案的排序值, 进而可得到方案的排序结果. 该方法具有概念清晰、计算简单和易于应用软件实现等特点, 为解决现实中基于多属性在线评价信息的方案排序问题提供了一种新途径. 需要指出的是, 本文的研究工作仅限于在线评价信息是多级评分的情形, 今后的研究工作需要考虑评价信息是文本形式的情形.

参考文献(**References**)

[1] Lee Y J, Hosanagar K, Tan Y. Do I follow my friends or the crowd? Information cascades in online movie ratings[J]. Management Science, 2014, 61(9): 2241-2258.

[2] Fang B, Ye Q, Kucukusta D, et al. Analysis of the perceived value of online tourism reviews: Influence of readability and reviewer characteristics[J]. Tourism Management, 2016, 52(1): 498-506.

[3] Najmi E, Hashmi K, Malik Z, et al. CAPRA: A comprehensive approach to product ranking using customer reviews[J]. Computing, 2015, 97(8): 1-25.

[4] Zhang K, Narayanan R, Choudhary A. Voice of the customers: mining online customer reviews for product feature-based ranking[C]. Proc of the 3rd Conf on Online Social Networks. Boston: Usenix, 2010: 1-9.

[5] Zhang K, Cheng Y, Liao W, et al. Mining millions of reviews: a technique to rank products based on importance of reviews[C]. Proc of the 13th Int Conf on Electronic Commerce. Amsterdam: Elsevier, 2011: 12-19.

[6] Saaty T L. The analytic hierarchy process[M]. New York: McGraw-Hill, 1980: 43-48.

[7] Hwang C L, Yoon K. Multiple attribute decision making: Methods and applications[M]. Berlin: Springer, 1981: 106-108.

[8] Wang Y M, Luo Y. Integration of correlations with standard deviations for determining attribute weights in multiple attribute decision making[J]. Mathematical & Computer Modelling, 2010, 51(1): 1-12.

[9] 樊治平. 多属性决策的一种新方法[J]. 系统工程, 1994, 12(1): 15-17.
(Fan Z P. A new method for multiple attribute decision making[J]. Systems Engineering, 1994, 12(1): 15-17.)

[10] 尤天慧, 樊治平. 区间数多指标决策中确定指标权重的一种客观赋权法[J]. 中国管理科学, 2003, 11(2): 92-95.
(You T H, Fan Z P. An objective method for determining attribute weights in multiple attribute decision making with intervals[J]. Chinese J of Management Science, 2003, 11(2): 92-95.)

[11] Chen T Y, Li C H. Determining objective weights with intuitionistic fuzzy entropy measures: A comparative analysis[J]. Information Sciences, 2010, 180(21): 4207-4222.

[12] 王翯华, 朱建军, 方志耕. 基于灰色关联度的多阶段语言评价信息集结方法[J]. 控制与决策, 2013, 28(1): 109-114.
(Wang H H, Zhu J J, Fang Z G. Aggregation of multi-stage linguistic evaluation information based on grey incidence degree[J]. Control and Decision, 2013, 28(1): 109-114.)

[13] Xu X. A note on the subjective and objective integrated approach to determine attribute weights[J]. European J of Operational Research, 2004, 156(2): 530-532.

[14] Byun H S, Lee K H. A decision support system for the selection of a rapid prototyping process using the modified TOPSIS method[J]. Int J of Advanced Manufacturing Technology, 2005, 26(11/12): 1338-1347.

[15] Koski J. Defectiveness of weighting method in multicriterion optimization of structures[J]. Communications in Applied Numerical Methods, 1985, 1(6): 333-337.

[16] Peng Y, Kou G, Li J. A fuzzy promethee approach for mining customer reviews in Chinese[J]. Arabian J of Science and Engineering, 2014, 39(6): 5245-5252.

[17] Chen K, Kou G, Shang J, et al. Visualizing market structure through online product reviews: Integrate topic modeling, TOPSIS, and multi-dimensional scaling approaches[J]. Electronic Commerce Research and Applications, 2015, 14(1): 58-74.

[18] Zhang W, Xu H, Wan W. Weakness finder: Find product weakness from Chinese reviews by using aspects based sentiment analysis[J]. Expert Systems with Applications, 2012, 39(11): 10283-10291.

(责任编辑：孙艺红)