

基于层次分析模型的产品多属性综合排序

王巍, 赵铁军, 辛国栋, 徐冰

(哈尔滨工业大学计算机科学与技术学院, 黑龙江 哈尔滨 150001)

摘要:为了解决产品的多属性综合排序问题,提出了一种基于层次分析模型的建模方法。通过将决策问题按目标层、准则层和方案层等分解为不同的层次结构,然后求得每一层次各元素对上一层次某元素的优先权重,再用加权和方法归并各参评方案对总目标的最终权重,从而完成产品的综合排序任务。实验结果证明所提方法在产品综合排序任务上的有效性。

关键词:层次分析模型;图模型;比较关系;综合排序

中图分类号:TP391.2

文献标识码:A

Multiattribute ranking based on analytic hierarchy model

WANG Wei, ZHAO Tie-jun, XIN Guo-dong, XU Bing

(School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China)

Abstract: An approach based on analytic hierarchy model for solving product multiattribute ranking was proposed. The approach decomposed a decision problem into hierarchy of more easily comprehended sub-problems, such as goal level, criteria level, and alternative level. Then priorities among the elements of the hierarchy were established towards its higher hierarchy. Synthesize these priorities to yield a set of overall priorities for the alternative hierarchy by using weighted sums. The experiment results show the effectiveness of analytic hierarchy model.

Key words: analytic hierarchy model, graph model, comparative relation, comprehensive ranking

1 引言

近年来,自然语言处理领域的很多机构和学者开展了情感分析方面的研究,如情感信息抽取、情感极性识别等。情感信息抽取和情感极性识别是情感分析的基础任务,它们为情感分析的应用任务,如评价实体的排序、情感文摘等提供必要的支持。本文以基本的情感分析技术为基础,研究评价实体的排序任务,即针对同一领域的实体,考察这些实体之间的比较关系,生成实体的综合排序结果。

在情感分析领域,已有的研究主要针对产品评论进行实体的排序,相关的研究主要分为如下

2 种:基于单一属性的产品排序和产品综合排序。迄今为止,主流的排序算法都是围绕用户对产品属性的情感倾向来开展,如对于单一属性的产品排序,研究者们通过计算产品在某种属性上的极性分值,并对这些分值进行排序,获得产品的属性排序结果^[1,2];而对于产品综合排序,研究者们通过计算产品在全属性上的极性分值,获得产品的综合排序结果^[3,4]。然而通过对潜在用户购买需求的分析,发现仅使用单一属性来排序产品是不够的,因为大多数用户所关注的产品属性不止一种,而且不同用户所关注的产品属性也有所不同,如下面 3 种情形。

情形 1:甲用户关注汽车的性价比、油耗和

收稿日期:2015-07-08;修回日期:2015-12-20

基金项目:国家高技术研究发展计划(“863”计划)基金资助项目(No.2015AA015405);国家自然科学基金资助项目(No.61402134, No.61173073, No.61272384, No.61172099);国际科技合作基金资助项目(No.2014DFA11350)

Foundation Items: The National High Technology Research and Development Program of China (863 Program) (No.2015AA015405), The National Natural Science Foundation of China (No.61402134, No.61173073, No.61272384, No.61172099), International Science and Technology Cooperation of China (No.2014DFA11350)

[动力]。

情形 2：乙用户关注汽车的[发动机性能]、内部空间和外观。

情形 3：丙用户关注汽车的价格、内饰和售后服务。

从这 3 个用户的购买需求来看，除了“动力”属性（在情形 2 中表现为“发动机性能”）是甲乙 2 个用户共同关注的属性外，上述 3 个用户所关注的产品属性各不相同。如果仅考察产品在单一属性上的排序结果，则无法帮助这 3 个用户做出合理的购买决定。从另一个角度，如果考察产品的综合排序结果（所有的评价属性），由于这类排序算法将每一个产品属性看作为同等重要的特征，且该算法为每一个用户提供相同的产品排序列表。

为了将产品的属性排序和潜在用户的需求更好地融合起来，本文提出了一种建模方法，该方法通过构建层次分析模型来实现产品的多属性综合排序。具体地，将与决策相关的元素分解为目标、准则（和子准则）、方案等层次，目标层是模型的决策目标，即指产品的综合排序或择优；准则层（和子准则层）是模型的决策依据，指产品的用户定义属性（UDF, user-defined features）和自然属性（CF, crude features）；方案层指的是评价产品。本文可以将用户的需求转化为权值运用在层次分析模型中，类似地，产品的属性排名也可以转化为权值应用在模型中。对于产品的属性排序，本文通过建立图模型来实现。具体地，将每个评价产品作为图中的节点，并使用产品的比较关系将相关的产品连接起来并设置边的权值，然后对比较关系图进行产品分值的计算。本文采用了类似 PageRank^[4]的方法对图模型进行迭代计算，最终获得图中每个节点的分值，即每个产品的属性排名分值。将本文提出的方法应用在汽车领域的语料，实验结果表明本文方法能够更好地适应用户的购买需求，且方法的性能优于以往的综合排序方法。同时也证明了层次分析模型对于产品多属性综合排序的有效性。

2 相关工作

在情感分析领域，与实体排序相关的工作主要可以分为基于属性的实体排序和实体的综合排序。对于这 2 类排序任务而言，主要研究工作都是围绕

着产品属性的情感极性情况来开展。基于属性的实体排序仅考察实体在某种属性上的情感极性，并对这些极性进行情感计算，获得最终的排序列表。Popescu 和 Etzioni^[1]首先识别出产品的属性，然后对与属性相关的观点进行识别，进而确定观点的情感极性并对产品的相对质量进行评估。他们的工作仅使用了评价文本的直接观点，对实体之间的比较观点没有考虑。Zhang 等^[2]使用大量的用户评论数据实现了基于属性的产品排序，他们首先识别出某类产品的全部属性，对于每一种属性，使用含有该属性的直接观点和比较观点来构建产品的有向图模型，并对图模型进行挖掘来确定产品的相对质量。他们的工作综合考察了产品的直接观点和比较观点，所构建的图模型能够较合理地反映产品之间的关系。

对于实体的综合排序，大多数研究者平等看待每一种产品属性，即认为每种属性对于产品的综合排序都有相同的贡献，因而研究者们对产品每种属性上的情感极性求和，再根据这些极性分值进行产品的综合排序。McGlohon 等^[5]设计了多种基于统计和基于启发式的排序模型，并在产品评论、商家评论和电影评论上对这些排序模型进行评测，最终基于平均值的统计方法获得了最佳的性能。Kurashima 等^[4]首先识别实体之间的比较关系，从大量的比较关系生成一个有向图，该有向图用于建模潜在用户的购买行为，并评估每个实体相对于其他实体的重要性。Zhang 等^[3]将基于属性的排序方法应用在产品综合排序上，通过向用户展示产品综合排序结果和属性排序结果来帮助用户做出购买决策。Zhang 等^[6]还根据评论是否对用户有帮助，以及评论的发表时间等因素计算每篇评论的重要性，使用这些评论重要性权值计算产品的综合排序分值。他们没有识别产品属性及属性的情感极性，而是使用评论篇章的重要性分值来排序产品。Wang 等^[7]提出了一种概率回归模型，用于从评论的篇章评分来计算每个属性的评分以及属性的重要性权值，在实体排序时只使用那些具有相似喜好的用户评论，根据这些评论计算属性的评分及重要性，进而使用加权模型计算实体的综合评分。他们的工作考虑了属性的重要性，然而该重要性权值来源于整篇评论的评分，只能反映一部分已有用户的喜好，无法反映出潜在用户的需求。

3 基于层次分析的产品综合排序

3.1 层次分析模型

综合评价指的是运用多个评价指标对多个参评对象进行评价（排序或择优），使用多种评价指标计算出一个综合指标值，根据这个综合指标值对评价对象优劣程度进行排序。针对产品的综合评价问题，本文提出了一种建模方法，该方法通过构建层次分析模型来完成产品的多属性综合排序。层次分析模型将与综合排序相关的元素分解成目标层、准则层、方案层等多个层次，在此基础上进行定量的分析，层次分析模型如图 1 所示。

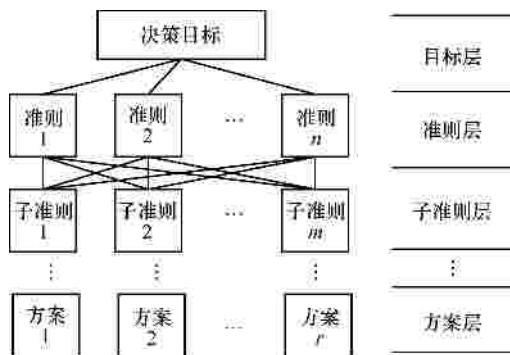


图 1 层次分析模型

图 1 的层次分析模型将决策问题按目标层、准则层、子准则层直至具体的参评方案层的顺序分解为不同的层次结构，然后求得每一层次各元素对上一层次某元素的优先权重，再用加权和方法归并各参评方案对总目标的最终权重，最终权重最大者即为最优方案。

3.2 系统框架概述

对产品的综合排序，本文建立了一个 4 层的层次分析模型，如图 1 所示。在该模型中，目标层是产品的决策目标，即产品的综合排序；准则层和子准则层是产品的评价指标，准则层是指产品的用户定义属性，子准则层是产品的自然属性；方案层是指参评产品。在这个模型中，方案层对子准则层某元素的权重通过构建有向图模型来计算；子准则层对准则层某元素的权重通过计算自然属性的出现频率得到；准则层对目标层的权重通过用户输入获得。系统总体框架如图 2 所示。

本文定义了 10 种 UDF 属性，如外观、动力和油耗等。每一个用户定义属性包含若干个 CF 属性，如外观包含车前、车后、尺寸、外形等自然属性。

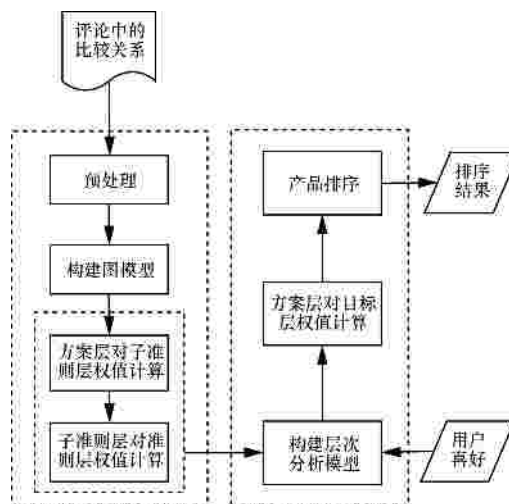


图 2 系统总体框架

3.2.1 图模型的构建

本文基于产品在属性上的比较关系来构建有向图，建立有向图模型的目的在于获取每一种产品关于属性的重要性分值（排名），并利用该分值计算出方案层对子准则层某元素的权值。在所构建的图模型 $G_f = (V, E)$ 中，每个节点代表一个产品，每条边代表 2 个产品之间的关系，建立起一个基于比较关系的有向图。在该有向图中， V_i 表示第 i 个节点； N 表示节点的总数； PC_{ij} 表示比较关系中 V_i 好于 V_j 的数量； NC_{ij} 表示比较关系中 V_i 不如 V_j 的数量； PS_i 表示与节点 V_i 相关的正面观点数； NS_i 表示与节点 V_i 相关的反面观点数。

在建立图模型时，如果节点 V_i 与 V_j 之间有比较关系，且 $PC_{ji} > NC_{ji}$ ，即在属性 f 上节点 V_j 优于 V_i ，则图中存在一条从 V_i 到 V_j 的有向边 E_{ij} ，这是因为潜在用户在购买产品时，总是倾向于从较差的产品向较好的产品转移。为了获得该有向边上的权值，定义 T_{ij} 为从节点 V_i 向节点 V_j 的转移概率，该概率的计算方法为

$$T_{ij} = \begin{cases} \frac{PC_{ji}}{PC_{ji} + NC_{ji}}, & i \neq j \\ \frac{PS_i}{PS_i + NS_i}, & i = j \end{cases} \quad (1)$$

在转移矩阵 T 中，由节点 V_i 向其相邻节点 V_j 转移的概率通过 PC 和 NC 来计算，自转移概率通过 PS 和 NS 来计算。然后由转移概率 T_{ij} 计算出边 E_{ij} 对应的权值 W_{ij} ，如式(2)所示。

$$W_{ij} = \frac{T_{ij}}{\sum_{k=1}^N T_{ik}} \quad (2)$$

为了求得每一个产品在产品集中的相对重要性，本文为每一个节点计算其重要性分值，节点 V_i 的分值由与 V_i 相邻的节点 V_j 来确定，节点重要性分值的计算方法为

$$R(V_i) = (1-d) \sum_{j=1}^N (W_{ji} R(V_j)) + d \quad (3)$$

其中， $R(V_i)$ 和 $R(V_j)$ 分别表示节点 V_i 和其相邻节点 V_j 的相对重要性， d 为跳转因子，跳转因子指用户停止访问 V_i 的相邻节点，转而随机访问图中任一节点的概率。经过多次迭代计算可以得到一个收敛的 $R(V_i)$ 值，这个值就是节点 V_i 的分值。

例如，有节点集合 $V = \{A, B, C, D\}$ ，评论文本中基于属性 f_k 的比较关系如下：

$PC_{AB} = 3, NC_{AB} = 1, PC_{AC} = 2, NC_{AC} = 1,$
 $PC_{AD} = 4, NC_{AD} = 1, PC_{BD} = 3, NC_{BD} = 2,$
 $PC_{DC} = 5, NC_{DC} = 2。$

基于以上关系，构建有向图模型如图 3 所示。在该图中有 A、B、C、D 这 4 种产品，节点上的数值表示该节点的自转移概率，有向边上的数值表示节点之间的转移概率，图中节点的转移关系都是从一个较差的产品转移至一个较好的产品。

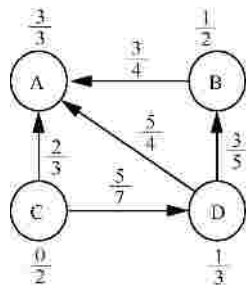


图 3 基于比较关系的图模型

例如，对于节点 A 而言，它有 3 个相邻节点 B、C、D，且在属性 f_k 上 A 超过了 B、C、D，因此 A 的自转移概率为 $\frac{3}{3}=1$ ；由产品之间的比较关系可以求得节点之间的转移概率。

3.2.2 方案层对子准则层的权值

在求得图模型中每一个节点关于属性 f_k 的重要性分值后，将该分值转化成方案层对子准则层某元素的权重，方案 i 关于子准则 k 的权重 $w_{ki}^{(4)}$ 计算如式(4)所示。

$$w_{ki}^{(4)} = \frac{R_{f_k}(V_i)}{\sum_{j=1}^N R_{f_k}(V_j)} \quad (4)$$

3.2.3 子准则层对准则层的权值

子准则是语料库中直接出现的自然属性，是与评论直接相关的，所以在赋予权值的时候应直接考虑该属性在对比语料中的出现次数，如“发动机”这一属性，评论中如果“发动机”出现的次数很多，说明有多数人对发动机这一性能非常在意，所以“发动机”的权值应该比较高。根据实际需求子准则 i 对于准则 k 的权重计算方法如式(5)所示。

$$w_{ki}^{(3)} = \begin{cases} \frac{C_{ki}}{\sum_{j=1}^N C_{kj}}, & i \in k \text{ 类别} \\ 0, & i \notin k \text{ 类别} \end{cases} \quad (5)$$

3.2.4 层次分析模型的构建

1) 层次分析模型的构建步骤

层次分析模型的基本思路与人对一个复杂的决策问题的思维、判断过程大体上是一致的。本文基于层次分析模型对产品的综合评价大致可以分为以下几个步骤。

将决策问题分解为 4 个层次：最上层为目标层，即选择或排序产品；最下层为方案层，有 P_1 、 P_2 、 P_3 等被比较的产品；中间层为准则层、子准则层，其中，准则层有外观、动力、油耗等 10 个评价准则，子准则层是评论中的自然属性。

确定各准则对于目标的权重。

通过建立图模型确定各方案对每一子准则的权重，及各子准则对于每个准则的权重。

将方案层对子准则层的权重、子准则层对准则层的权重及准则层对目标层的权重进行综合，最终确定方案层对目标层的权重。

基于以上产品综合评价信息，生成排序结果。汽车的层次分析模型如图 4 所示。

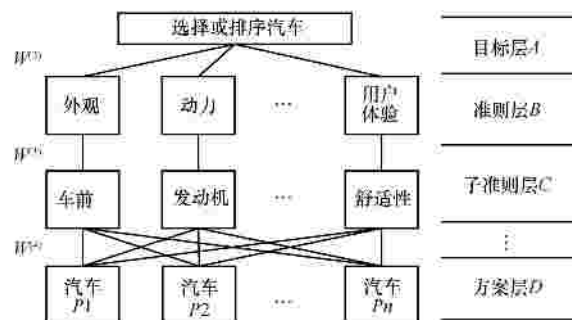


图 4 汽车的层次分析模型

2) 准则层对于目标层的权值计算

层次分析模型构建好之后，下一步是计算准则层对于目标层的权值，准则层对于目标层的权值应该和客户需求直接相关，如某客户关注“油耗”这一指标，“油耗”就应该被赋予更大的权值，所以准则层对于目标层的权值可以由用户输入给出，模型根据用户的需求（用户输入的权值）来计算最终的排序结果。根据用户对几种用户定义属性的排序，采用式(6)^[8]计算准则层对于目标层的权值（ K 个属性中计算 k 属性值）。

$$W_k^{(2)} = \frac{K+1-k}{\sum_{i=1}^K i} \quad (6)$$

假设有 3 个属性，那么这些属性各自的权值是 $[\frac{3}{6}, \frac{2}{6}, \frac{1}{6}]$ 。

3) 方案层对于目标层的权值计算

层次分析模型的工作过程大致如下：设准则层 B 对决策层 A 的权向量为 $w^{(2)} = (w_1^{(2)}, w_2^{(2)}, \dots, w_n^{(2)})^T$ ，其中， n 为准则的数量；子准则层 C 对于准则层 B_k 的权向量 $w_k^{(3)} = (w_{k1}^{(3)}, w_{k2}^{(3)}, \dots, w_{km}^{(3)})^T$ ，其中 $k=1, 2, \dots, n, m$ 为子准则的数量，以 $w_k^{(3)}$ 为列向量构成矩阵 $W^{(3)} = (w_1^{(3)}, w_2^{(3)}, \dots, w_n^{(3)})$ ；方案层 D 对于子准则层 C_k 的权向量 $w_k^{(4)} = (w_{k1}^{(4)}, w_{k2}^{(4)}, \dots, w_{kr}^{(4)})^T$ ，其中 $k=1, 2, \dots, m, r$ 为方案的数量，以 $w_k^{(4)}$ 为列向量构成矩阵 $W^{(4)} = (w_1^{(4)}, w_2^{(4)}, \dots, w_m^{(4)})$ ；最后方案层 D 对决策层 A 的权向量为

$$w^{(5)} = W^{(4)}W^{(3)}w^{(2)} \quad (7)$$

对 $w^{(5)}$ 的权值进行排序，获得的排序列表即为产品的综合排序结果，其最大值即为最佳决策。

具体而言，首先从语料中抽取与第 k 个自然属性相关的比较关系，基于这些比较关系建立有向图模型，对建立的有向图进行迭代计算，得到所有产品（方案）在属性 k （子准则）上的重要性分值 $V_k = (v_{k1}, v_{k2}, \dots, v_{kr})$ ，将这些分值利用式(4)进行处理得到 $w_k^{(4)}$ 。进而由于子准则的出现频率计算 $w_k^{(3)}$ ，根据用户的输入来计算 $w_k^{(2)}$ 。最终由式(7)计算方案层对目标层的权值。

4 实验结果与分析

4.1 实验设置

4.1.1 实验数据

本文使用中文情感分析评测 (COAE, Chinese opinion analysis evaluation) 2012 的评测数据作为实验数据。其中选用了任务 2 的汽车领域的语料，一共 813 个比较句中的 1 100 个比较关系，涉及到 208 种汽车实体。本文将这些比较关系中的比较实体作为图模型的节点，使用图模型计算某种自然属性上各产品的排序权值，进而使用层次分析模型对产品进行综合排序。

4.1.2 评价方法

本文使用 Spearman 相关系数和信息检索领域的平均准确率 (MAP, mean average precision) 对算法的实体排序结果进行评价。Spearman 相关系数 r 的定义为

$$r(S_a, S_b) = 1 - \frac{6 \sum_{i=1}^n (S_{ai} - S_{bi})^2}{n(n^2 - 1)} \quad (8)$$

其中， S_a 和 S_b 分别代表随机变量 S_a 和 S_b 的权值向量（同为升序或降序）， r 为这 2 个变量的相关系数。 r 越接近 1，2 个变量越相关； r 越接近 0，2 个变量越不同。

平均准确率：单个主题的平均准确率是每个相关实体排序后准确率的平均值。定义为

$$MAP = \frac{\sum_{i=1}^N P_i(r)}{N} \quad (9)$$

其中， N 代表单个主题相关实体的个数， $P_i(r)$ 指查全率为 r 时的第 i 个实体的查准率。

4.1.3 对比实验系统

为了对本文提出的基于层次分析模型的算法进行更好的评价，以及与前人的方法进行对比，设计了如下 3 个对比实验系统。

Baseline：产品集合中所有的产品被随机指定排序的顺序。这种对比实验系统是一种随机排序的方法。

图结构算法：该对比实验系统主要参照 Kurashima 等^[4]的方法，不考虑各自然属性之间的差别，建立基于比较关系的图模型对产品进行综合排序。

本文算法：首先基于比较关系在每种自然属性上构建图模型，获得方案层对每个子准则的权值；进而基于自然属性的频率计算子准则对每个准则的权值；为了方便与其他系统的比较，在此系统中，设置各准则对目标的权值均为 1，即不考虑用户对各 UDF 属性的需求。

4.2 实验结果比较

为了获得更合理准确的实验结果，本文将产品集中所有产品按照价格区间分成 5 个类别，这 5 个类别分别是 5 万以下、5~10 万、10~20 万、20~50 万、50 万以上；基于这些价格区间，从“汽车之家”论坛上获取各区间汽车产品的口碑排名，作为标准的排序结果；对于每一个价格区间，计算 2 个对比实验系统的排序结果与标准排序结果之间的 Spearman 相关系数。对于每个价格区间的产品，分别请 2 名汽车领域的专业人士给出他们的排序结果，将每个价格区间中各对比实验系统的排序结果和 2 名专家的排序结果计算 MAP 值。

当求出各个价格区间的 Spearman 相关系数和 MAP 值后，对这些相关系数和 MAP 值取平均，可以获得最终的 Spearman 相关系数和 MAP 值，结果如表 1 所示。从实验结果可以看出，本文的方法好于 2 种对比实验系统。证明本文方法中所建立的层次分析模型和基于比较关系的属性排序图模型对于产品的综合排序任务是有效的。

表 1 产品排序的实验结果对比

方法	Spearman 相关系数	MAP@10 (专家 1)	MAP@10 (专家 2)
Baseline	0.454 7	0.434 3	0.423 7
图结构算法	0.493 6	0.604 2	0.593 6
本文算法	0.614 4	0.728 4	0.692 5

从表 1 中可以看出，Baseline 系统的性能最差，其原因主要在于这种方法仅依靠随机的方法对产品进行排序，该方法每次产生的排序结果不同，且既没有使用评论中现有用户的产品比较信息，也没有考虑潜在用户的需求。与之相比，图结构算法使用了评论中的比较关系，并基于比较关系构建有向图模型来计算每个产品的综合重要性，基于这些重要性分值对产品进行排序，该方法增加了现有用户的产品比较信息，使排序任务有据可依，排序后的 Spearman 相关系数和 MAP 值都得到了显著的提升。

尽管图结构算法使用了现有用户对产品的比

较信息，该方法还存在着一定的局限性。在该方法中，一方面没有考虑各个自然属性的差别，将所有自然属性看作同等重要的属性来构建图模型并进行产品的重要性排序；另一方面该方法没有考虑潜在用户的购买需求，因而对所有用户提供相同的购买建议。

为了利用现有用户对产品比较关系的信息，同时考虑潜在用户的购买需求，本文构建了层次分析模型，该方案通过产品的比较关系构建基于属性的图模型，计算每一种产品（方案）相对于每种自然属性（子准则）的权值，并根据自然属性的出现频率计算每一种自然属性（子准则）相对于每种 UDF 属性（准则）的权值。进而接收用户对 UDF 属性的权值（为了方便比较，此处都设置为 1），由这 3 个层次的权值计算出方案层对目标的权值，完成产品的多属性综合排序。该方案融合了以往用户的比较观点和潜在用户的购买需求，与标准排序结果的 Spearman 相关系数和不同专家的排序结果的 MAP 值均有显著提升，证明了层次分析模型对产品综合排序的有效性。

5 结束语

本文针对产品的多属性综合排序问题提出了一种基于层次分析模型的建模方法。该方法将与综合排序相关的元素分解成目标层、准则层、方案层等多个层次，在此基础上进行定量的分析。通过计算每一个层次对上一层次的权重，并将多个层次的权重综合起来求得每个产品的最终排序权重，来实现产品的多属性综合排序。实验结果证明本文提出的方法在产品的多属性综合排序任务是有效的。

参考文献：

- [1] POPESCU A M, Etzioni O. Extracting product features and opinions from reviews[C]//Natural language Processing and Text Mining. Springer London, c2007: 9-28.
- [2] ZHANG K, NARAYANAN R, CHOUDHARY A. Voice of the customers: mining online customer reviews for product feature-based ranking[C]//The 3rd Conference on Online Social Networks. c2010.
- [3] ZHANG K, NARAYANAN R, CHOUDHARY A. Mining online customer reviews for ranking products[R]. EECS department, Northwestern University, 2009.
- [4] KURASHIMA T, BESSHO K, TODA H, et al. Ranking entities using comparative relations[C]//Database and Expert Systems Applications.

Springer Berlin Heidelberg, c2008: 124-133.

- [5] MCGLOHON M, GLANCE N S, REITER Z. Star quality: quality: aggregating reviews to rank products and merchants[C]//ICWSM. Washington DC, USA, c2010: 114-121.
- [6] ZHANG K, CHENG Y, LIAO W, et al. Mining millions of reviews: a technique to rank products based on importance of reviews[C]//The 13th International Conference on Electronic Commerce. Liverpool, UK ,ACM, c2011: 12-19.
- [7] WANG H, LU Y, ZHAI C. Latent aspect rating analysis on review text data: a rating regression approach[C]//The 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Phuket, Thailand, ACM, c2010: 783-792.
- [8] GOODWIN P, WRIGHT G. Decision analysis for management judgement[M]. Wiley: Chichester, UK, 2004.



赵铁军 (1962-), 男, 黑龙江哈尔滨人, 哈尔滨工业大学教授、博士生导师, 主要研究方向为自然语言处理和人工智能应用。



辛国栋 (1976-), 男, 山东青岛人, 哈尔滨工业大学讲师, 主要研究方向为社交网络和网络安全。

作者简介:



王巍 (1975-), 女, 黑龙江齐齐哈尔人, 哈尔滨工业大学博士生, 主要研究方向为自然语言处理、情感分析和多文档文摘。



徐冰 (1975-), 女, 黑龙江大庆人, 哈尔滨工业大学副教授, 主要研究方向为自然语言处理和情感分析。