[2010(1)]

中国企业运筹学

China Journal of Enterprise Operations Research

中国运筹学会
企业运筹学分会 主编

电子科技大学出版社

[2018年]

中国企业运筹学

China Journal of Enterprise Operations Research

[2010(1)]

中国企业运筹学

China Journal of Enterprise Operations Research

中国运筹学会
企业运筹学分会 主编

电子科技大学出版社

图书在版编目（CIP）数据

中国企业运筹学 / 中国运筹学会企业运筹学分会编.
—成都：电子科技大学出版社，2010.7
ISBN 978-7-5647-0563-3

Ⅰ.①中… Ⅱ.①中… Ⅲ.①运筹学—应用—企业管理—中国—高等学校—教材 Ⅳ.①F279.23

中国版本图书馆 CIP 数据核字（2010）第 135589 号

中国企业运筹学

中国运筹学会
企业运筹学分会 主编

出　　版：电子科技大学出版社（成都市一环路东一段 159 号电子信息产业大厦　邮编：610051）
策划编辑：郭　庆
责任编辑：谢应成
主　　页：www.uestcp.com.cn
电子邮箱：uestcp@uestcp.com.cn
发　　行：新华书店经销
印　　刷：成都火炬印务有限公司
成品尺寸：185mm×260mm　　印张　13.25　　字数　425　千字
版　　次：2010 年 7 月第一版
印　　次：2010 年 7 月第一次印刷
书　　号：ISBN 978-7-5647-0563-3
定　　价：50.00 元

目　录

管理科学

工业工程

投融资理论

物流供应链管理

市 场 营 销

其　他

面向缺失数据的客户价值区分集成模型研究

肖 进 贺昌政

（四川大学工商管理学院 四川 成都 610064）

摘 要：当客户数据包含缺失时，已有的大多数研究工作采用“两步式”客户价值区分策略。将集成学习技术与数据分组处理（group method of data handling, GMDH）理论相结合，提出了面向缺失数据的客户价值区分“一步式”集成模型 GDCEMV。实证分析表明，与已有的 4 种“两步式”策略以及 1 种“一步式”策略相比，GDCEMV 在客户数据包含缺失情况下能够取得更好的客户价值区分效果。

关键词：客户价值区分；缺失数据；“一步式”集成模型；数据分组处理；集成学习

中图分类号：F830.91 **文献标识码**：A

0 引言

客户价值区分是根据客户为企业创造价值的能力对客户进行分类，提供针对性的产品、服务和营销模式的过程，能够使企业更合理的分配资源，有效地降低成本，同时获得更有利可图的市场渗透[1]。客户信用评估、客户流失预测、客户交叉销售等均属于客户价值区分的范畴，它们都是按价值的不同维度对客户进行区分，比如客户信用评估是按客户当前价值的大小进行分类，而客户流失预测则按客户潜在价值的大小来分类。

在客户价值区分中，客户数据在很多时候都包含缺失数据。如在利用调查问卷搜集数据时，由于客户不愿回答一些敏感的问题（如收入、年龄），或没有足够的知识和经验回答某些问题，都有可能造成数据缺失。Yim 等学者[2]通过问卷调查搜集数据来研究顾客满意度问题，他们指出在回收的 450 份问卷中有 90 份含有大量的缺失数据。不仅是来自问卷的数据，很多来自企业的 CRM 数据库的客户资料数据也常常包含缺失。以全球 500 强企业霍尼韦尔（Honeywell）为例，虽然他们的数据收集有一套严格的规范，但该公司的客户数据库中数据缺失率仍然高达 50%[3]！

目前常用的客户价值区分模型如神经网络、支持向量机等都要求模型的训练集数据必须是完整的[4]，只要待分类客户有一个特征的值是缺失的，模型就没有办法对其进行分类。为了解决这一问题，Kim 等学者提出了一个客户价值区分的 3 阶段框架[5]：1）数据收集和预处理；2）客户价值分类建模；3）营销策略的制定。在这一框架下，当客户数据包含缺失时，大量研究是先对缺失数据进行预处理，使数据集变得完整，然后在完整的数据集上建立客户价值区分模型，这两个阶段是彼此相对独立进行的，我们称之为“两步式”客户价值区分策略[6]。其中使用得最多的预处理方法还是插补法，包括单值插补方法如均值替代、线性回归插补、EM 插补和多重插补方法[7]。如 Lessmann and Voß[8]提出了一种基于支持向量机的层次参考模型用于信用评估，在建立模型之前采用均值替代来处理缺失数据，Shao 等人[9]在进行客户流失预测时，根据属性是连续还是离散而采用不同的预处理方案，对于连续属性，采用均值替代；对于离散的属性，则将缺失值看做该变量的一种新的状态。然而，插补法仍有不足之处，常用的插补方法都是基于随机缺失假设的，且都需要假定数据服从某一分布模型，但在实际应用中，各种缺失方式经常是交织在一起的，采用的假设、模型不合理，将影响后继分类器的学习效果[10]。因此，“两步式”策略有待改进。

为了弥补插补法的不足，近年来在数据挖掘领域，有学者尝试使用集成学习技术直接构建面向缺失数据的分类模型。如 Krause 等人[11]提出了一种直接为含有缺失的数据进行分类的集成方法 learn^{++} MF（Learn^{++} for

基金项目：国家自然科学基金（70771067），中德国际合作项目（70911130228）

Missing Features)，简称 LMF。该方法首先从特征空间选择一系列特征子集，通过映射计算得到若干训练子集，并在每个训练子集上产生一个弱的基本分类器构成基分类器池，然后对于每一个待分类样本 x^*（可能含有缺失特征），用基分类器池中没有使用 x^*中的缺失特征的那些基本分类器为其分类，最后将这些基分类器的分类结果进行投票得到最终分类结果。LMF 属于“一步式”客户价值区分的范畴。实验分析表明，LMF 具有较好的分类性能。然而，需要指出的是，该方法着重研究了测试集中包含缺失的情形，没有考虑训练集中也包含缺失的情形，但在现实的客户分类问题中，训练集和测试集通常都存在数据缺失，同时，Krause 等人也指出，当某一测试样本包含较多缺失特征时，很可能在基分类器池中一个可行的分类器都找不到，这时 LMF 方法将无法对其进行分类[11]。最后，对于每一个待分类样本 x^*，LMF 方法将基分类器池中满足要求的全部基分类器的分类结果通过多数投票来得到最终分类结果，但这些基分类器之间可能存在冗余，因此若能从中选择一个适当的子集进行集成，将有望进一步提高分类性能[12]。

本文在 Krause 等人提出的 LMF 基础上，提出了一种直接面向缺失数据的客户价值区分“一步式”集成策略 GDCEMV（GMDH Based Dynamic Classifier Ensemble for Missing Value）。该方法能够充分利用数据集中已知的样本信息，在进行客户分类之前不需要事先对缺失数据进行处理，从而减少了对数据缺失机制假设以及数据分布模型的依赖。此外 GDCEMV 也能处理训练集和测试集中同时存在数据缺失的客户价值区分问题。

1 常用缺失数据处理方法

目前，处理缺失数据的主要方法有个案删除法、单值插补法和多重插补法。

1.1 个案删除法

个案删除法（Listwise Deletion，LS）是最简单的处理缺失数据的方法。在这种方法中如果任何一个变量含有缺失，就把相应的个案从数据集中删除。如果缺失值所占比例比较小的话，这一方法十分有效。然而，当缺失数据所占比例较大，特别是当缺失数据是非随机分布的时候，这种方法可能导致数据发生偏离，从而得出错误的结论[3]。同时，当数据集中每个样本均包含一定的缺失时，此方法是不可用的。

1.2 单值插补法

（1）均值替换法

采用均值替换法（Mean Substitution，MS）处理缺失数据时，通常将变量的属性分为数值型和非数值型来分别进行处理。如果缺失值是数值型的，就根据该变量在其他所有对象的取值的平均值来填充该缺失的变量值；如果缺失值是非数值型的，就根据统计学中的众数原理，用该变量在其他所有对象的取值次数最多的值来补齐该缺失的变量值。均值替换法也是一种简便、快速的缺失数据处理方法。使用 MS 方法插补缺失数据，对该变量的均值估计不会产生影响。但这种方法是建立在完全随机缺失（MCAR）的假设之上的，而且会造成变量的方差和标准差变小。

（2）EM 方法

EM 算法[13]是根据所得观测数据，获得对模型参数估计的一种方法，其核心思想就是根据已有的数据来递归估计似然函数。EM 算法包括两步：E 步指根据 D_{obs} 和第 t 次迭代得到的模型参数 $\theta^{(t)}$ 来预测 $D_{mis}^{(t)}$；M 步是指根据 D_{obs} 和 $D_{mis}^{(t)}$ 来估计 $\theta^{(t+1)}$。给定模型参数 θ 的初值 $\theta^{(0)}$，重复 E 步和 M 步，直到参数估计收敛为止，收敛时得到的 $D_{mis}^{(t)}$ 可看做插补值。作为一种迭代方法，EM 算法最主要的缺点在于计算成本较大，还可能出现局部收敛和伪收敛的现象。在小样本、高缺失率且有很多参数需要估计时，有可能发生唯一似然最大值不存在的情况。此外，EM 算法要求提前给出分布形式，这在大多数实际问题中都是难以做到的。

（3）回归插补法

回归插补法（Rregression Imputation, RI）首先需要选择若干个预测缺失值的自变量，然后建立回归方程估计缺失值，即用缺失数据的条件期望值对缺失值进行替换。与前述几种插补方法比较，该方法利用了数据库中

尽量多的信息，但该方法也有很多不足：1）它虽然是一个无偏估计，但是却容易忽视随机误差，低估标准差和其他未知性质的测量值，而且这一问题会随着缺失信息的增多而变得越严重；2）研究者必须假设存在缺失值所在的变量与其他变量存在线性关系，但很多时候这种关系是不存在的。

1.3 多重替代法

单值插补法容易造成数据分布的扭曲，可能会使得到的结果偏差较大。为改善这一弊端，美国哈佛大学统计学家 Rubin 提出了多重插补方法（Multiple Imputation, MI）。[7]首先，MI 用一系列可能的值来替换每一个缺失值，以反映被替换的缺失数据的不确定性。然后，用标准的统计分析过程对多次替换后产生的若干个数据集进行分析。最后，把来自于各个数据集的统计结果进行综合，得到总体参数的估计值。

2 “一步式”集成模型

本研究提出的“一步式”集成模型 GDCEMV 中的集成包含两层含义：第一，该模型将 Kim 等学者[5]提出三阶段客户价值区分框架中的第 1 阶段的数据预处理与第 2 阶段的建立客户分类模型进行集成；第二，将多分类器集成方法引入到客户价值分类建模中。

设分类问题包含 n 个属性，其训练集和测试集分别是 D_{train} 和 D_{test}，它们可能都含有一定比例的缺失值。GDCEMV 方法的基本思想如下：首先将训练集 D_{train} 按缺失特征的个数分成 n 个子集：$D_1, D_2, \cdots, D_n$，分别表示缺失 1 个，2 个，…，n–1 个，0 个特征的样本构成的集合，然后在每个子集上根据特征的缺失比例的不同赋予一个不同的被抽样权重，在此基础上随机选择若干特征子集，通过映射计算得到一系列训练子集，将训练子集中含有缺失的样本删除，并训练得到一系列基本分类器。进一步的，对于每一个测试样本 x^*，在每个子数据集中寻找到 x^*的 K 个近邻组成 x^*的局部区域 L_{ocal}，然后利用基于 GMDH 的动态分类器集成选择方法[14]找到在 L_{ocal} 中具有最大分类精度的一个集成 E^*，并用 E^*来为 x^*进行分类，这样每个数据子集 D_i（i=1, 2, …, n）中都得到了 x^*的一个分类结果，最后采用投票法得到最终分类结果。GDCEMV 方法的伪代码如下：

输入：训练集 D_{train} 和测试集 D_{test}、近邻数 K 以及在每个子数据集上选择的特征子集个数 F；

输出：测试集中每个样本 x^*的最终分类结果。

1．根据缺失属性的个数，将训练集 D_{train} 分成 $D_1, D_2, \cdots, D_n$；

2．对于每个 D_i, i=1, 2, …, n,如果 D_i 不为空，则

2.1 统计每个特征上缺失的样本个数，设分别为 $m_1, m_2, \cdots, m_n$，若 $m_i \neq 0$，则为第 i 个特征赋予被抽样权重 $w_i=1/m_i$,否则令 $w_i=2$，最后将 $w_1, w_2, \cdots, w_n$ 进行标准化处理得到 $w'=(w_1', w_2', \cdots, w_n')$。

2.2 根据 w'随机抽取 F 个特征子集，且每个特征子集的特征个数为原始属性个数的一半，通过映射计算得到 F 个训练子集 $S_1, S_2, \cdots, S_F$ 。

2.3 删除 S_j （j=1, 2, …, F）中含有缺失的样本，并在每个 S_j 上训练一个基本分类器。

3．对每个测试样本 x^*，

3.1 在每个 D_i（D_i 不为空）中，寻找 x^*的 K 个近邻构成局部区域 L_{ocal}，然后利用基于 GMDH 的集成选择方法从 D_i 中的 F 个基本分类器中找到在 L_{ocal} 上具有最高分类精度的一个集成 E^*，并用 E^*为 x^*进行分类，得到分类结果；

3.2 将在所有 D_i 上得到的 x^*的分类结果进行多数投票，得到 x^*的最终分类结果。

3 实证分析

3.1 试验设置

为了分析本文提出的面向缺失数据的“一步式”集成模型 GDCEMV 的分类性能，以国际标准数据库（UCI）[15]中的澳大利亚客户信用评估数据集“Australia”为例进行实证分析。该数据集共有 690 个客户样本，包含两类 {*good*,*bad*}，分别表示信用好和信用差的客户。其中，信用差的客户样本 383 人，信用好的客户样本 307 人，

两者的比例为 1.2476，属于类别分布相对比较均衡的数据集。为了保护商业机密，公开的数据对属性名和属性值做了符号代换，共有 14 个属性，其中，定量属性 6 个，定性属性 8 个。同时，在该数据集中，有 37 个客户样本包含 1 个或者多个缺失值，数据缺失率 5.36%。此数据集也来自于 UCI 数据库，本身不包含噪声[15]。因此，使用 GDCEMV 来对其进行分类是比较合理的。

为便于处理，本研究统一将数据集中的连续型属性按照 Fayyad and Irani [16]提出的方法进行预离散化。而为了得到训练集和测试集，随机地将整个数据集分成 3 等份，其中 1/3 的样本用于测试，而余下 2/3 的样本构成训练集。同时，将支持向量机（support vector machine，SVM）作为基本分类模型。在 GDCEMV 算法中，有两个重要参数：近邻数 K 以及在每个子数据集上选择的特征子集个数 F。通过反复试验，我们取 K=3，F=50。最后，每一个分类结果均是取 10 次实验的平均值。

我们将 GDCEMV 的客户价值区分效果与现有的面向缺失数据的“两步式”分类策略如先采用 MS 进行插补再建立单一 SVM 模型、先采用 EM 方法进行插补再建立单一 SVM 模型、先采用 RI 进行插补再建立单一 SVM 模型、先采用 MI 进行插补再建立单一 SVM 模型（依次简记为 MS-SVM，EM-SVM，RI-SVM，MI-SVM），以及 Krause 等人提出的“一步式”客户价值区分策略 LMF 的客户价值区分效果进行了比较，如表 1 所示。

表 1　客户信用评估矩阵

样本状态	预测为负类	预测为正类	合计
负类（信用好的类）	A	B	$A+B$
正类（信用差的类）	C	D	$C+D$
合计	$A+C$	$B+D$	$A+B+C+D$

注：A 代表实际为负类，预测为负类的样本个数；B 代表实际为负类，预测为正类的样本个数；C 代表实际为正类，预测为负类的样本个数；D 代表实际为正类，预测为正类的样本个数。

为了评价各种方法的信用评估效果，本文引入如表 1 所示的评价矩阵。在这一基础上，我们引入 4 个常用的模型评价准则：①总的准确率 Accuracy=（$A+D$）/（$A+B+C+D$）×100%；　②ROC 曲线及 AUC 值。在类别不平衡问题上仅仅用总的准确率这一个指标显然是不够的。ROC 曲线是常用的类别分布不平衡条件下分类模型的评价准则。针对两类问题的 ROC 图是一个真正率—伪正率图，其中真正率= D/（$C+D$）×100%，伪正率= B/（$A+B$）×100%，X 轴用伪正率表示，Y 轴用真正率表示。由于直接比较不同分类模型的 ROC 曲线很不方便，因此，很多时候，人们使用 ROC 曲线下方的面积，即 AUC（Area Under the ROC Curve）值来比较分类器性能的优劣。③少数类准确率（又叫覆盖率）：Minor accuracy= D/（$C+D$）×100%；④多数类准确率：Major accuracy= A/（$A+B$）×100%。

3.2　试验结果分析

图 1 给出了 6 种策略在“australia”数据集上的 ROC 曲线，其中 ROC 曲线与 x 轴围成的区域面积越大，则表明对应的方法的整体分类性能越好。从图 1 可以看出，6 种分类策略的 ROC 曲线之间都存在一定程度的交叉，但总的来说，在绝大多数时候 GDCEMV 的 ROC 曲线均位于最上方，其次是 EM-SVM 策略和 LMF 策略的 ROC 曲线，处于最下面的是 MS-SVM 策略的 ROC 曲线。因此，我们可以大致判断，在“Australia”数据集上，GDCEMV 策略具有最好的分类性能，而 MS-SVM 策略的分类性能是最差的。

表 2 给出了 6 种策略在“Australia”数据集上的 AUC 值、总的准确率、信用好的客户准确率以及信用差的客户准确率。表中的粗体表示每一列对应的最大值。从表中可以看出，GDCEMV 策略具有最高的 AUC 值、总的准确率以及信用差的客户准确率，由此，我们得出结论，在“Australia”数据集上，本文提出的面向缺失数据的“一步式”集成策略 GDECMV 具有最好的信用评估效果。进一步需要指出的是，在 6 种分类策略总的准确率差别不是太大的情况下，只有 GDCEMV 策略的信用差的客户准确率是高于信用好的客户准确率的。如 LMF 策略，虽然总的准确率仅次于 GDECMV 策略，但主要是由于它对信用好的客户准确率比较高（在 6 种策略中是最高的），而对信用差的客户准确率只有 0.8370，因此，该策略的信用评估性能并不令人满意。

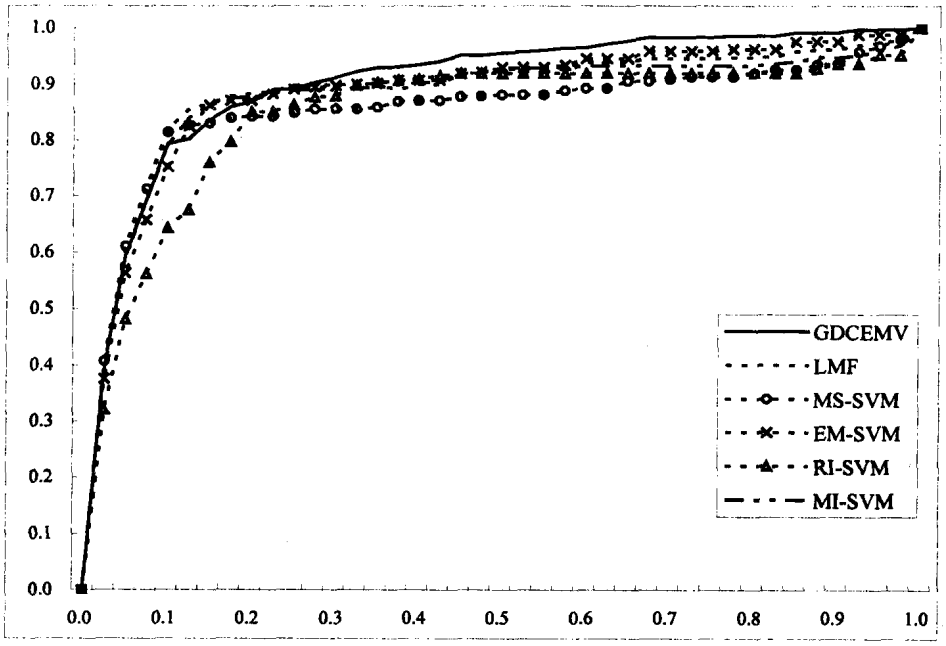

图 1　6 种策略在 australia 数据集上的 ROC 曲线

表 2　6 种策略在 australia 数据集上的分类性能比较

Method	AUC	总的准确率	信用好的客户准确率	信用差的客户准确率
GDCEMV	**0.9184**	**0.8766**	0.8653	**0.8919**
LMF	0.8998	0.8682	**0.8913**	0.8370
MS-SVM	0.8790	0.8473	0.8718	0.8143
EM-SVM	0.9024	0.8607	0.8736	0.8432
RI-SVM	0.8842	0.8502	0.8858	0.8024
MI-SVM	0.8935	0.8565	0.8804	0.8243

4　结论

本文主要关注缺失数据条件下的客户价值区分问题，针对已有的“两步式”价值区分策略的不足，提出了基于 GMDH 的“一步式”集成策略 GDCEMV。该策略不仅能够处理训练集中包含缺失的情形，也能处理训练集和测试集中同时包含缺失的情形，它能够在不对缺失数据进行插补的条件下，充分利用数据集中的未缺失的数据信息，从而减少了对数据缺失机制假设以及数据分布模型的依赖。在澳大利亚客户信用评估数据集“australia”上的实证结果表明，本文所提出的“一步式”集成策略具有很好的处理缺失数据的能力，与已有的 4 种“两步式”策略以及 1 种“一步式”策略相比，GDCEMV 能够取得更好的客户价值区分效果。

参考文献

[1] Venkatesan R, Kumar V. A customer lifetime value approach for customer selection and resource allocation strategy [J]. Journal of Marketing, 2004, 68（4）: 106-125

[2] Yim C K, Tse D K, Chan K W. Strengthening customer loyalty through intimacy and passion: roles of customer-firm affection and customer-staff relationships in services [J]. Journal of Marketing Research, 2008, 45: 741-756

[3] Lakshminarayan K, Harp S A, Samad T. Imputation of missing data in industrial databases [J]. Applied

Intelligence, 1999, 11（3）: 259-275

[4] Kim J, Hwang K J, Bae J K. Prediction of personal credit rates with incomplete data sets using cognitivemapping [C]. IEEE Computer Society Washington, 2007, 1912-1917

[5] Kim S Y, Jung T S, Suh, E H, et al. Customer segmentation and strategy development based on customer lifetime value: a case study [J]. Expert systems with Applications, 2006, 31（1）: 101-107

[6] Avery R B, Calem P S, Canner G B. Consumer credit scoring: do situational circumstances matter[J]? Journal of Banking and Finance, 2004, 28（4）: 835-856

[7] Rubin D B. Multiple Imputations for Nonresponse in Surveys [M]. New York: John Wiley and Sons, 1987

[8] Lessmann S, Voß S. A reference model for customer-centric data mining with support vector machines [J]. European Journal of Operational Research, 2009, 199（2）: 520-530

[9] Shao J B, Li X, Liu W H. The application of AdaBoost in customer churn prediction [C]. International Conference on Service Systems and Service Management, 2007, 1-6

[10] Jiang K, Chen H, Yuan S. Classification for incomplete data using classifier ensembles [C]. International Conference on Neural Networks and Brain, ICNN&B '05, 2005, 559-563

[11] Mohammed H S, Stepenosky N, Polikar R. An ensemble technique to handle missing data from sensors [C]. IEEE Sensors Applications Symposium Houston, Texas USA, 2006, 101-105

[12] Zhou Z H, Wu J, Tang W. Ensembling neural networks: many could be better than all [J]. Artificial Intelligence, 2002, 137（1-2）: 239-263

[13] Dempster A P, Laird N M, Rubin D B. Maximum likelihood from incomplete data via the EM algorithm [J]. Journal of the Royal Statistical Society, 1977, 39: 1-38

[14] Xiao J, He C Z. Dynamic classifier ensemble selection based on GMDH. in: Proceeding of the Second International Joint Conference on Computational Sciences and Optimization, 2009, 731-734

[15] Merz C, Murphy P. UCI repository of machine learning databases. 1995, http://www. ics. uci. edu/ ~mlearn/MLRepository. html

[16] Fayyad U M, Irani K B. Multi-interval discretization of continuous-valued attributes for classification learning [C]. Proceedings of 13th International Joint Conference on Artificial Intelligence, 1993, 1022-1027

Customer Value Differentiation Ensemble Model for Missing Values

XIAO Jin HE Chang-zheng

（Business School, Sichuan University, Chengdu 610064, China）

Abstract: When the customer data contain missing values, most existing researches adopt "two-step" customer value differentiation （CVD） strategies. This paper combines ensemble learning with group method of data handling, and presents "one-step" CVD model—GDCEMV for missing values. The empirical results show that GDCEMV can get better effects compared with the existing 4 "two-step" CVD strategies and a "one-step" strategy.

Keywords: customer value differentiation; missing value; "one-step" ensemble model; GMDH; ensemble learning

基于理想关联度法的旅行社专线产品竞争力评价

周文坤

（上海大学管理学院　上海　200444）

摘　要：对旅行社专线产品模式的概念和类型进行了深入的分析，运用可持续发展的科学观设计旅行社专线产品竞争力模型和评价指标体系，利用理想关联度方法对旅行社专线产品的总体状态进行定量综合评价，科学地反映旅游市场上竞争态势。以四川省内游旅行社专线产品进行实证分析，得出了科学合理的结论。

关键词：旅行社专线产品；多指标决策；评价；竞争力；关联系数

中图分类号：C812　　**文献标识码**：A

0　引言

竞争是市场经济条件下组织或产业生存和发展的内在要求与动力，随着社会经济的发展与经济全球化、一体化进程的加快，无论一个国家、一个地区还是一个组织、一个产业，都必须面对激烈的市场竞争的挑战，而组织或产业要在激烈的竞争中获胜，竞争力是关键性的决定因素。竞争力是指在竞争性市场中一个企业或组织所具有的能够持续地比其他企业或组织更有效地向市场提供产品或服务，并获得赢利和自身发展的综合素质[1,2]。其实质是与竞争对手比较而产生的相对位置关系，其结局是优胜劣汰。因此，对竞争力的研究成为经济和管理界广泛关注和高度重视的课题，并且将一直成为具有无限生命力的研究领域。

旅游市场指由旅游产品引起的在供需双方交换过程中所反映出来的各种经济行为和经济关系的总和[1]。旅游业是由提供各种能满足旅客需求的产品或服务的行业所构成的集合，其中包括旅游观赏娱乐业（旅游景区/点）、餐饮住宿业（宾馆、酒店）、旅行社、交通通信业、旅游购物业等。这些产业是以旅游景区/点为核心，通过旅行社为桥梁和纽带，协调旅游业与各相关行业的企业之间的组织工作与依存关系，连接旅游业的各个组成部分，共同构成旅游相关产业聚集。旅行社通过从相关企业采购所需的产品和服务来满足旅客的各种需求，形成旅游组合产品销售给旅客。旅行专线产品作为旅游产业中的一种重要表现形式，在拉动和满足旅游市场需求方面起着非常关键的作用。因为旅行社与旅客直接发生接触关系，比较了解旅客的需求和偏好，能够有的放矢地设计和安排适合游客的旅游产品，有效地组织游客的旅游活动顺序，最大限度地满足旅客的体验需求，同时在旅游场所的其他企业采购相关的产品和服务。旅行社和顾客正是通过旅游市场这个媒介进行交易，而旅行社专线产品成为重要的交易载体，因而它不仅是旅行社向顾客提供产品和服务的基础，而且也是旅行社同行之间开展市场竞争的重要工具与主要形式[3-12]（如图 1 所示）。

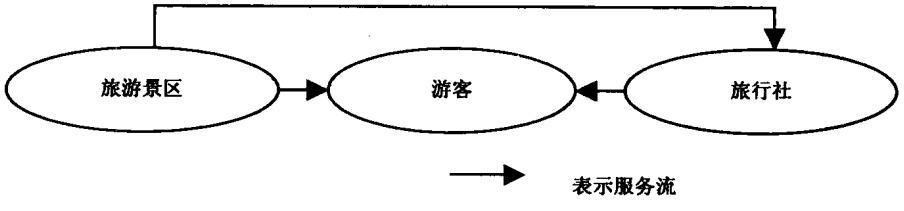

图 1　旅行社专线产品结构模型

基金项目：上海市教委科研创新项目（08YS19）

根据旅客流动方向，可将旅行社专线产品分为国内和境外旅行社专线产品。而前者又可以分为出省旅行社专线产品和入省旅行社专线产品，后者又可以分为出境旅行社专线产品和入境旅行社专线产品（如图 2 所示）。

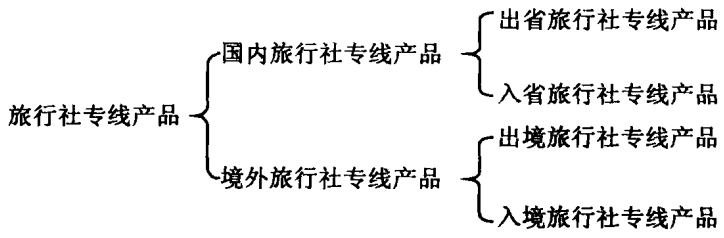

图 2　旅行社专线产品的类型

1　旅行社专线产品竞争力评价指标体系设计

旅行社专线产品竞争力是其综合实力的反映，是以资源与管理为基础的外在表现形式。因而，在评价其竞争力时必须全面、客观、动态地从整体的视角进行考察，以防止片面、主观、静态地评价所带来的不良影响，进而缺乏可信性与权威性。

1.1　指标体系设计的基本原则

旅行社专线产品竞争力评价指标体系不是一些指标的简单堆积和随意组合，而是根据某些基本的原则建立起来的，并且能够反映旅行社专线产品竞争态力状况的指标集合。为了科学、合理地进行综合评价，应遵循以下基本原则：

（1）科学性与实用性原则

指标体系应充分反映和体现旅行社专线产品竞争力的内涵，要有合理的层次结构，能客观综合地反映旅行社专线产品的经济效益，体现旅行社专线产品在竞争力方面的核心指标。

（2）系统性与层次性原则

评价指标体系是一个复杂的系统，它可以分解若干个子系统。应在不同层次上采用不同的指标，有利于政府、企业决策者在不同层次上进行调控，对旅游资源进行有效地配置，提升旅行社专线产品的竞争力。

（3）动态性与稳定性原则

指标体系中的指标内容在一定的时期内应保持相对稳定，就可以比较和分析旅行社专线产品竞争力发展的过程并预测其发展趋势。同时，由于旅行社专线产品竞争力培育是一个持续渐进的发展过程，所以设计指标体系时应充分考虑系统的动态变化，能综合反映企业现状和发展趋势。

（4）可测性与可比性原则

评价指标体系应尽可能量化，对于一些难以量化的指标，其影响重大的也可以用定性指标来描述。同时，这些指标的计算方法应当明确，不要过于复杂，计算所需要的数据也应比较容易获得和比较可靠。

（5）完备性与简明性原则

指标体系要内容简单、明了与准确，并具有代表性。指标体系中的指标数量不宜过大，在相对比较完备的情况下，指标的数目应尽可能地压缩便于操作。

1.2　指标体系设计的路径与内容

竞争力是企业或产业以具有的能够持续地合理控制和充分运用各种资源，在与竞争对手的角逐中建立在经济、环境和社会责任等有形资产与无形资产基础上的竞争优势能力总和。旅行社专线产品竞争力是通过其在市场销售产品或服务而反映出来的生产力，最终由市场表现来衡量和检验。可构建评价指标体系，如图 3 所示。

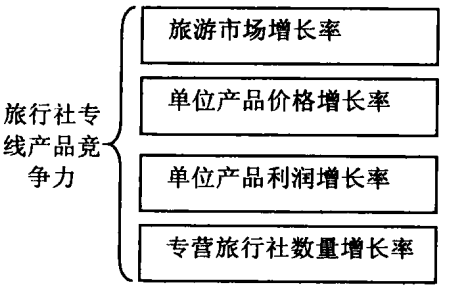

图 3　旅行社专线产品竞争力评价体系

旅行社专线产品的竞争力是外在的综合表现形式，而其决定因素是文化内涵与有效的管理，如图 4 所示。

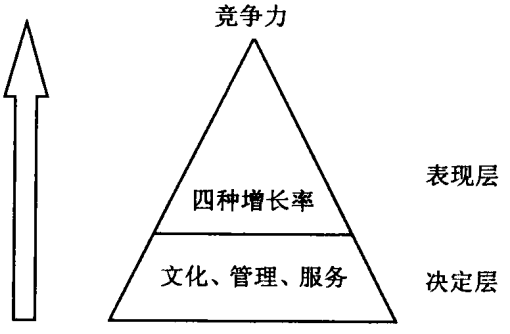

图 4　旅行社专线产品竞争力模型

2　评价旅行社专线产品竞争力的理想关联度法

若多指标决策模型由指标和决策方案两个要素组成，其中由 m 个方案组成决策方案集 $X=\{x_1,x_2,\cdots,x_m\}$，由 n 个评价方案的指标组成指标集 $U=\{u_1,u_2,\cdots,u_n\}$。对于方案 $x_i \in X$，按第 j 个指标 u_j 进行测度，得到 x_i 关于 u_j 的指标值为 a_{ij}　（$i=1,2,\cdots,m$；　$j=1,2,\cdots,n$），　$A=(a_{ij})_{m\times n}$。

在指标集 U 中有效益型与成本型两种类型指标。为了消除各指标的不同物理量纲对决策结果的影响，对决策矩阵 A 中各元素进行规范化处理。设 I_j　（$j=1,2$）分别表示效益型与成本型指标的下标集，用下列规范决策矩阵的计算公式，将决策矩阵 A 转化为规范化矩阵 $R=(r_{ij})_{m\times n}$：

$$r_{ij}=\begin{cases} a_{ij}/\sum_{i=1}^{m}a_{ij} & j\in I_1, i\in M \\ (a_{ij})^{-1}/\sum_{i=1}^{m}(a_{ij})^{-1} & j\in I_2, i\in M \end{cases} \tag{1}$$

在规范化矩阵 $R=(r_{ij})_{m\times n}$ 中，对指标 u_j 而言各决策方案指标值的标准差和平均差分别为

$$S_j(w)=\sqrt{\frac{1}{m}\sum_{i=1}^{m}(\| r_{ij}w_j-\frac{1}{m}\sum_{k=1}^{m}r_{kj}w_j\|)^2}=w_j\sqrt{\frac{1}{m}\sum_{i=1}^{m}l^2(r_{ij},r_j)}\quad(j=1,2,\cdots,n) \tag{2}$$

$$V_j(w)=\frac{1}{m}\sum_{i=1}^{m}\| r_{ij}w_j-\frac{1}{m}\sum_{k=1}^{m}r_{kj}w_j\|=w_j\frac{1}{m}\sum_{i=1}^{m}l(r_{ij},r_j)\quad(j=1,2,\cdots,n) \tag{3}$$

其中，$r_j=\frac{1}{m}\sum_{k=1}^{m}r_{kj}$，表示指标 u_j 下各方案的平均指标值；$l(r_{ij},r_j)$ 表示指标 u_j 下的平均指标值 r_j 与方案 x_i 的指标值 r_{ij} 的相离度。

根据的基本思想，权重向量 w 的选择应该使在各指标下所有决策方案的组合差之和最大。因此，构造目标函数[13-14]

$$\min D(w)=\sum_{j=1}^{n}(\alpha S_j(w)+\beta V_j(w))=\sum_{j=1}^{n}w_j[\alpha\sqrt{\frac{1}{m}\sum_{i=1}^{m}l^2(r_{ij},r_j)}+\beta\frac{1}{m}\sum_{i=1}^{m}l(r_{ij},r_j)]$$

$$\alpha+\beta=1,\quad \alpha\geqslant 0,\quad \beta\geqslant 0 \tag{4}$$

其中，α 和β 体现了决策者对各种信息的偏好程度。

$$\text{记}\ \sigma_j=\sqrt{\frac{1}{m}\sum_{i=1}^{m}l^2(r_{ij},r_j)},\quad \delta_j=\frac{1}{m}\sum_{i=1}^{m}l(r_{ij},r_j) \tag{5}$$

因而，求解 w 等价于求解下面的单指标非线性规划问题：

$$\min D(w)=\sum_{j=1}^{n}w_j(\alpha\delta_j+\beta\xi_j) \tag{6}$$

$$\text{s. t.}\ \sum_{j=1}^{n}w_j^2=1,\quad w_j\geqslant 0 \tag{7}$$

解该模型得

$$w_j=\frac{\alpha\delta_j+\beta\xi_j}{\sqrt{\sum_{j=1}^{n}(\alpha\delta_j+\beta\xi_j)^2}} \tag{8}$$

对权重进行归一化处理，得到

$$w_j^*=\frac{\alpha\delta_j+\beta\xi_j}{\sum_{j=1}^{n}(\alpha\delta_j+\beta\xi_j)} \tag{9}$$

在求出指标的最优权重向量 $w=(w_1^*,\ w_2^*,\ \cdots,\ w_n^*)^{\mathrm{T}}$ 后，再结合规范化矩阵 $R=(r_{ij})_{m\times n}$，可以得到加权矩阵 $\overline{R}=(w_j^*r_{ij})_{m\times n}$。

在规范化矩阵 $R=(r_{ij})_{m\times n}$ 中，方案 x_i 对应的向量为 $\tilde{z_i}=(r_{i1},r_{i2},\cdots,r_{in})$，$i\in M$，定义方案 x_i 关于指标 u_j 的关联系数为

$$\xi_{ij}=\frac{\min\limits_i\min\limits_j\{L_{ij}\}+\rho\max\limits_i\max\limits_j\{L_{ij}\}}{L_{ij}+\rho\max\limits_i\max\limits_j\{L_{ij}\}}\quad(i=1,2,\cdots,m,\ \ j=1,2,\cdots,n) \tag{10}$$

其中，$L_{ij}=|r_{ij}-r_j|$ 为两者的距离，$r_j=\max\limits_{1\leqslant i\leqslant m}\{r_{ij}\}$，$\rho$ 为分辨系数，$\rho\in[0,\ 1]$，一般取值为 0.5，方案与理想最优方案 x_0 的关联系数向量为 $\xi_i=(\xi_{i1},\ \xi_{i2},\ \cdots,\ \xi_{in})^{\mathrm{T}}$。由此，各方案与理想方案的关联系数矩阵为

$$\xi=\begin{bmatrix}\xi_{11} & \xi_{12} & \cdots & \xi_{1n}\\ \xi_{21} & \xi_{22} & \cdots & \xi_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ \xi_{m1} & \xi_{m2} & \cdots & \xi_{mn}\end{bmatrix} \tag{11}$$

因此，方案 x_i 与 x_0 的综合关联度为

$$\zeta_i=\sum_{j=1}^{n}w_j^*\xi_{ij},\quad i=1,2,\cdots,m \tag{12}$$

显然，ζ_i 越大，表示被评价方案 x_i 与理想方案 x_0 越接近，因而决策方案 x_i 越好。各方案按综合关联度 ζ_i 大小对方案进行排序，即可得到决策方案集合中的最优方案。

3 实证分析

以四川省内游旅行社专线产品数据为例，应用理想关联度方法进行实证分析，如表 1 所示。

表 1　四川省内游旅行社专线产品的数据分析（2000-2005 年）

指标 / 专线产品	旅游市场增长率（I1）	单位产品价格增长率（I2）	单位产品利润增长率（I3）	专营旅行社数量增长率（I4）
A 九寨—黄龙	0.28	-0.53	-0.07	0.26
B 峨眉—乐山	0.15	-0.38	0.02	1.16
C 四姑娘山	0.32	0.33	-0.10	0.71
D 都江堰—青城山	0.10	-0.45	-0.06	0.18
E 蜀南竹海	0.08	0.06	-0.07	0.55
F 西岭雪山—花水湾	-0.09	-0.05	-0.22	0.42
G 雅安碧峰峡	0.24	-3.57	-0.07	0.43
H 海螺沟	1.35	0.09	-0.06	0.35
I 广元剑门关	0.07	0.11	0.10	0.58
J 邓小平故居	0.33	-0.03	-0.07	0.70
K 西昌邛海—泸沽湖	0.02	0.07	-0.18	0.81
L 三星堆	0.34	0.13	0.01	0.35

数据来源：根据四川旅游政务网（http：//www. scta. gov. cn/web/）及相关资料整理。

（1）评价决策矩阵 A 和规范化矩阵 R 分别如下：

$$A=\begin{bmatrix} 0.28 & -0.53 & -0.07 & 0.26 \\ 0.15 & -0.38 & 0.02 & 1.16 \\ 0.32 & 0.33 & -0.10 & 0.71 \\ 0.10 & -0.45 & -0.06 & 0.18 \\ 0.08 & 0.06 & -0.07 & 0.55 \\ -0.09 & -0.05 & -0.22 & 0.42 \\ 0.24 & -3.57 & -0.07 & 0.43 \\ 1.35 & 0.09 & -0.06 & 0.35 \\ 0.07 & 0.11 & 0.10 & 0.58 \\ 0.33 & -0.03 & -0.07 & 0.70 \\ 0.02 & 0.07 & -0.18 & 0.81 \\ 0.34 & 0.13 & 0.01 & 0.35 \end{bmatrix} \quad R=\begin{bmatrix} 0.0878 & 0.1256 & 0.0909 & 0.1359 \\ 0.0470 & 0.0900 & -0.0260 & 0.0305 \\ 0.1003 & -0.0782 & 0.1299 & 0.0498 \\ 0.0313 & 0.10665 & 0.0779 & 0.1963 \\ 0.0251 & -0.0142 & 0.0909 & 0.0643 \\ -0.0282 & 0.0118 & 0.2857 & 0.0841 \\ 0.0752 & 0.8460 & 0.0909 & 0.0822 \\ 0.4232 & -0.0213 & 0.0779 & 0.1010 \\ 0.0219 & -0.0261 & -0.1299 & 0.0609 \\ 0.1034 & 0.0071 & 0.0909 & 0.0505 \\ 0.0063 & -0.0166 & 0.2338 & 0.0436 \\ 0.1066 & -0.0308. & -0.0130 & 0.1010 \end{bmatrix}$$

（2）应用上文公式（9）计算得到各指标权重

σ_1=0. 1105；σ_2=0. 2374；σ_3=0. 1052；σ_4=0. 0443；δ_1=0. 0674；δ_2=0. 1392；δ_3=0. 07161；δ_4=0. 0336

假定两种决策信息地位相同，则取$\alpha=\beta=1/2$ 时，有

w_1^*=0. 2198；w_2^*=0. 4654；w_3^*=0. 2185；w_4^*=0. 0963

（3）对应的各旅行社专线关于理想方案的关联度值为

ζ（A）=0. 5623；ζ（B）=0. 4976；ζ（C）=0. 5339；ζ（D）=0. 5749；ζ（E）=0. 5204；ζ（F）=0. 5862；ζ（G）=0. 6761；ζ（H）=0. 6300；ζ（I）=0. 4795；ζ（J）=0. 5297；ζ（K）=0. 5550；ζ（L）=0. 5220

（4）根据关联度由大到小排序，有 G>H >F>D >A >K >C> J> L>E>B >I。

雅安碧峰峡的综合关联度值处于第一位，关键是其善于不断地寻找“卖点”刺激公众，具有丰富的概念想象空间，在以自然景观为主的四川省旅游市场上，独树一帜，具有明显的产品特色，总体效果不错。九寨—黄龙专线是四川省内游旅专线的主力品牌，资源号召力强，产品知名度和美誉度都较高，在旅游市场上保持强劲的增长势头。而其他各专线产品的关联度值也具有很强的代表性。因而，各旅行社专线产品关联度度值较好地反映他们各自的总体地位和发展状态。