

使用机器学习算法基于行为决定因素的宫颈癌早期检测

王文盈

云南财经大学统计与数学学院, 云南 昆明

收稿日期: 2023年9月17日; 录用日期: 2023年10月8日; 发布日期: 2023年10月23日

摘要

宫颈癌是最常见的女性生殖道恶性肿瘤, 发病率在女性恶性肿瘤中居第二位, 在某些发展中国家甚至位居首位。在我国, 宫颈癌也是危害女性健康与生命的重要疾病。幸运的是, 女性可以通过接种HIV疫苗和定期筛查来预防这种疾病, 但目前预防方法的结果和参与度都比较低。一方面是由于经济条件的限制, 另一方面是因为女性本身预防知识和意识的缺乏。因此, 为了积极预防宫颈癌, 本文分别利用逻辑斯蒂回归、线性判别分析、二次判别分析、朴素贝叶斯、 K -近邻方法检测基于行为决定因素的宫颈癌风险, 并选择准确率最高的预测方法。而在正式进行预测分类前, 本文利用主成分分析对数据进行降维处理; 利用交叉验证对变量进行选择。

关键词

逻辑斯蒂回归, 线性判别分析, 二次判别分析, 朴素贝叶斯, K -近邻, 主成分分析, 交叉验证

Early Detection of Cervical Cancer Based on Behavioral Determinants Using Machine Learning Algorithms

Wenyang Wang

School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming Yunnan

Received: Sep. 17th, 2023; accepted: Oct. 8th, 2023; published: Oct. 23rd, 2023

Abstract

Cervical cancer is the most common malignancy of the female reproductive tract, ranking second in

incidence and even first in some developing countries. In China, cervical cancer is also an important disease that endangers women's health and life. Fortunately, women can prevent the disease through HIV vaccination and regular screening, but current prevention methods have low outcomes and participation. On the one hand, it is due to the restrictions of economic conditions, and on the other hand, it is because of the lack of prevention knowledge and awareness of women themselves. Therefore, in order to actively prevent cervical cancer, this paper used logistic regression, linear discriminant analysis, quadratic discriminant analysis, naive Bayes and K -nearest neighbor methods to detect the risk of cervical cancer based on behavioral determinants, and selected the prediction method with the highest accuracy. Before the formal prediction classification, this paper uses principal component analysis to reduce the dimension of the data. Use cross-validation to select variables.

Keywords

Logistic Regression, Linear Discriminant Analysis, Quadratic Discriminant Analysis, Naive Bayes, K -Nearest Neighbor, Principal Component Analysis, Cross Verification

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 研究背景

宫颈癌是威胁全球妇女健康及生命的主要恶性肿瘤之一，其发病率占女性恶性肿瘤第4位。尽管随着社会经济和医疗事业的发展，宫颈癌的发病率在发达国家已大幅下降，但在发展中国家宫颈癌的发病率与死亡率仍居高不下。在我国，宫颈癌新发病例达13.15万，死亡人数每年约为5.3万，其发病率和死亡率存在明显的地理差异[1]。

目前，控制宫颈癌发病率主要以预防为主。通过注射HIV疫苗、对适龄妇女人群进行宫颈癌筛查可以有效降低宫颈癌的发病率及其导致的死亡率。由于疫苗接种也不能提供全面的保护，其长期疗效尚未确定，因此通过了解个人健康行为对宫颈癌进行预防就变得很重要[2]。

机器学习已广泛用于疾病的早期检测[3][4]，甚至用于某些重大疾病的预测。一些来自分类技术的机器学习算法，如朴素贝叶斯、逻辑斯蒂回归、 K -NN、和支持向量机是数据挖掘中热门的算法。

在本文中，逻辑斯蒂回归、线性判别分析、二次判别分析、朴素贝叶斯和 K -NN分别被用作分类器，来检测基于行为及其决定因素的宫颈癌风险[5][6]，并将它们的检测结果进行对比，选择准确率最高的分类器。

2. 相关理论与方法

2.1. 主成分分析

主成分分析是利用少数几个综合原始数据集中大部分变异信息的典型变量概括数据集，是一种常见的降维方法。提取的主成分要尽可能多地包含原始信息，且互不相关，主成分的个数可以根据累计贡献率的阈值和碎石图来决定。

$$Z_1 = \phi_{11}X_1 + \phi_{21}X_2 + \cdots + \phi_{p1}X_p \quad (1)$$

$\phi_1 = (\phi_{11}, \phi_{21}, \dots, \phi_{p1})'$ 是第一主成分载荷向量，是提取原始变量信息最多的主成分。同理，可以定义第一主成分，是在剩余信息中保留变异信息最多的主成分，其方向与第一主成分垂直。

2.2. 交叉验证

统计建模往往分析的是 $n \gg p$ 的数据，而对于特征比较多的数据，一方面可以使用降维方法降低维度；另一方面，可以使用子集选择的方法进行变量选择，本文使用交叉验证方法来选择变量，核心算法见表 1。

Table 1. Cross validation algorithm
表 1. 交叉验证算法

| |
|-----------------------------------|
| 产生 k 各预测模型 $M_k, k=0,1,2,\dots$ |
| 计算每个模型的预测误差，并选择误差最小的模型 |

2.3. 逻辑回归

逻辑回归的本质是分类，通过构造 *Sigmoid Function* 来计算 Y 在每一类的概率，哪类概率大， Y 就被判到哪类。其缺点是，当维度 p 比较大时，算法很难收敛。

$$P(Y=1|X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \quad (2)$$

2.4. 生成式模型

1) 线性判别分析

线性判别分析是一种有监督的学习，当特征变量来自同方差的分布时，其判别分数与 X 是线性关系。

$$\delta_k(x) = x' \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k' \Sigma^{-1} \mu_k + \log \pi_k \quad (3)$$

2) 二次判别分析

与线性判别分析不同的是，二次判别分析主要用于特征之间具有非线性关系时的分类，此时特征变量是来自不同方差的分布。

$$\delta_k(x) = -\frac{1}{2} x' \Sigma_k^{-1} x + x' \Sigma_k^{-1} \mu_k - \frac{1}{2} \mu_k' \Sigma_k^{-1} \mu_k - \frac{1}{2} \log |\Sigma_k| + \log \pi_k \quad (4)$$

3) 贝叶斯判别分析

贝叶斯分析相较于其他判别方法有更强的假设，假设定量变量服从高斯分布，其缺点是没有对变量之间的关系进行研究。

$$p_k(x) = P(Y=k|X=x) = \frac{\pi_k f_k(x)}{\sum_{i=1}^K \pi_i f_i(x)} \quad (5)$$

2.5. K-近邻

K -近邻采用寻找“邻居”的方法找到类与类之间的分界线。 K 值越大，模型越简单，分界线越光滑。因此，在建模过程中需要多尝试几个 K 值(注： K 表示总邻居的个数)。

$$P(Y=j|X=x_0) = \frac{1}{K} \sum_{i \in N_0} I(y_i = j) \quad (6)$$

3. 数据说明及基本分析

3.1. 数据说明

本文选择的数据集将行为的决定因素：感知、意图、动机、主观规范、态度、社会支持和授权转化

为问卷，并将问卷分发给 72 名受访者。其中，有 22 名受访者是宫颈癌患者，50 名受访者不是宫颈癌患者。受访者均来自印度尼西亚雅加达的城市公民(注：本数据集不存在空值，同时符合分类算法的数据要求，因此不需要进行预处理)。数据来源于 UCI 机器学习数据库，具体变量情况见表 2。

Table 2. Table of data variable

表 2. 数据变量表

| 数据特征指标体系 | | | | |
|----------|-----------|----------------------------|----------------------------------|----------|
| 一级指标 | 二级指标 | 三级指标 | 原始变量名称 | 修改后的变量名称 |
| 个人行为 | 主观规范 | 性行为风险 | behavior_sexualRisk | X1 |
| | | 饮食行为 | behavior_eating | X2 |
| | | 个人卫生行为 | behavior_personalHygine | X3 |
| 决定因素 | 意图 | 意图聚合 | intention_aggregation | X4 |
| | | 意向承诺 | intention_commitment | X5 |
| | 态度 | 态度一致性 | attitude_consistency | X6 |
| | | 态度自发性 | attitude_spontaneity | X7 |
| | 主观规范 | 规范重要人物 | norm_significantPerson | X8 |
| | | 规范履行 | norm_fulfillment | X9 |
| 感知 | 感知漏洞 | perception_vulnerability | X10 | |
| | 感知严重度 | perception_severity | X11 | |
| | 动机 | 动机强度 | motivation_strength | X12 |
| 动机意愿 | | motivation_willingness | X13 | |
| 社会支持 | 社会支持 - 情感 | socialSupport_emotionality | X14 | |
| | 社会支持 - 欣赏 | socialSupport_appreciation | X15 | |
| | 社会支持 - 仪器 | socialSupport_instrumental | X16 | |
| 授权 | 授权 - 知识 | empowerment_knowledge | X17 | |
| | 授权 - 能力 | empowerment_abilities | X18 | |
| | 授权 - 愿望 | empowerment_desires | X19 | |
| | 分类变量 | | ca_cervix 1 = 有宫颈癌, 0 = 没有宫颈癌 | Y |

3.2. 描述性分析

3.2.1. 特征基本关系

由于部分分类算法对特征之间的关系进行了假设，因此本文计算了所有原始变量的相关系数，发现变量 X13~X19 之间存在一定的线性关系，相关系数均在 0.5 以上(见表 3)。即对于女性患者，社会在物质上给予客观支持，在情感上给予主观支持，有利于减弱女性患者的病耻感[7] [8]，帮助她们更加积极的面对应对，降低宫颈癌死亡率；而对于女性非患者，社会支持则有利于增强她们的医疗健康意识，从而积极主动去做定期筛查，降低宫颈癌发病率。

Table 3. Table of raw variable partial correlation coefficient**表 3.** 原始变量部分相关系数表

| | X13 | X14 | X15 | X16 | X17 | X18 | X19 |
|-----|--------|--------|--------|--------|--------|--------|--------|
| X13 | 1.0000 | 0.7017 | 0.6328 | 0.5129 | 0.5950 | 0.6181 | 0.5235 |
| X14 | 0.7017 | 1.0000 | 0.7411 | 0.6063 | 0.6341 | 0.7793 | 0.6265 |
| X15 | 0.6328 | 0.7411 | 1.0000 | 0.7371 | 0.5082 | 0.6302 | 0.6460 |
| X16 | 0.5129 | 0.6063 | 0.7371 | 1.0000 | 0.6264 | 0.6307 | 0.7109 |
| X17 | 0.5950 | 0.6341 | 0.5082 | 0.6264 | 1.0000 | 0.8365 | 0.7319 |
| X18 | 0.6181 | 0.7793 | 0.6302 | 0.6307 | 0.8365 | 1.0000 | 0.7436 |
| X19 | 0.5235 | 0.6265 | 0.6460 | 0.7109 | 0.7319 | 0.7436 | 1.0000 |

3.2.2. 各特征分布分析

为了分析数据的联合分布和边际分布形态, 本文利用 Shapiro-Wilk 统计量对宫颈癌数据联合正态性以及各变量的边际正态性进行了检验(见表 4), 发现数据的联合分布和边际分布均违背了正态性假设, 因此我们后续对其进行了降维处理。

Table 4. Multivariate normal distribution and marginal normal distribution test**表 4.** 多元正态分布及边际正态分布检验

| — | Shapiro 统计量 | P 值 | 正态分布(Yes/No) |
|---------|-------------|--------|--------------|
| 联合分布正态性 | 0.4687 | 0.0000 | No |
| X1 | 0.3188 | 0.0000 | No |
| X2 | 0.8408 | 0.0000 | No |
| X3 | 0.9389 | 0.0017 | No |
| X4 | 0.7527 | 0.0000 | No |
| X5 | 0.7295 | 0.0000 | No |
| X6 | 0.9362 | 0.0012 | No |
| X7 | 0.8206 | 0.0000 | No |
| X8 | 0.7297 | 0.0000 | No |
| X9 | 0.8197 | 0.0000 | No |
| X10 | 0.8988 | 0.0000 | No |
| X11 | 0.7817 | 0.0000 | No |
| X12 | 0.7570 | 0.0000 | No |
| X13 | 0.8971 | 0.0000 | No |
| X14 | 0.8772 | 0.0000 | No |
| X15 | 0.8878 | 0.0000 | No |
| X16 | 0.8622 | 0.0000 | No |
| X17 | 0.8424 | 0.0000 | No |
| X18 | 0.9092 | 0.0001 | No |
| X19 | 0.8472 | 0.0000 | No |

3.3. 数据预处理

为了防止数据过度训练, 本文通过随机抽样将数据分为训练集和测试机。同时, 由于本文选择的数据集变量众多(19 个), 且部分变量之间具有相关性, 因此在正式做预测分类前, 本文利用主成分分析对数据进行降维处理; 利用交叉验证对数据进行变量选择, 并将降维和子集选择的数据作预测分类后的结果进行对比。

3.3.1. 主成分降维

为了保留原始变量绝大部分信息, 根据累计方差贡献在 80%以上, 同时又最大程度降低特征维度, 本文最终选取了 7 个主成分, 并得出了每个主成分在原始变量的载荷矩阵(见表 5)和和其关于原始变量的线性组合。由于本文使用主成分降维只是一个中间过程, 并不是做最终的结果模型, 因此无需对主成分进行解释。

Table 5. Principal component load coefficient

表 5. 主成分载荷系数

| | PC1 | PC2 | PC3 | PC4 | PC5 | PC6 | PC7 |
|-----|---------|---------|---------|---------|---------|---------|---------|
| X1 | -0.1033 | -0.0656 | 0.2505 | -0.1501 | 0.5437 | -0.2304 | 0.5893 |
| X2 | -0.0019 | 0.0165 | -0.5120 | -0.1960 | -0.2074 | -0.5791 | -0.0382 |
| X3 | -0.2185 | -0.1794 | -0.2784 | 0.2919 | -0.2375 | -0.1599 | 0.2489 |
| X4 | -0.1271 | -0.0770 | -0.3351 | 0.4875 | 0.1838 | -0.1802 | 0.0472 |
| X5 | -0.0591 | 0.0024 | -0.3634 | -0.0196 | 0.6260 | 0.0678 | -0.1679 |
| X6 | 0.0291 | -0.1612 | -0.2899 | -0.2267 | -0.2629 | 0.4105 | 0.6320 |
| X7 | 0.0136 | 0.0833 | -0.4030 | -0.4972 | 0.1031 | 0.1977 | -0.1226 |
| X8 | 0.0182 | -0.4284 | -0.0154 | 0.0362 | -0.0098 | -0.1318 | -0.0780 |
| X9 | 0.0042 | -0.4920 | 0.0667 | -0.0719 | 0.0224 | -0.0474 | -0.0368 |
| X10 | -0.0580 | -0.4462 | 0.1198 | -0.2457 | 0.0133 | -0.0785 | -0.1561 |
| X11 | -0.0573 | -0.4791 | 0.0729 | -0.1026 | -0.0514 | 0.0645 | -0.1560 |
| X12 | -0.1945 | -0.1281 | -0.1719 | 0.3110 | 0.0828 | 0.5448 | -0.1392 |
| X13 | -0.3451 | -0.0402 | 0.0699 | 0.1278 | 0.0830 | -0.0274 | 0.1548 |
| X14 | -0.3640 | 0.0570 | 0.1080 | 0.0732 | -0.1329 | -0.0381 | -0.1207 |
| X15 | -0.3433 | 0.1114 | 0.1090 | 0.0049 | -0.1864 | -0.0148 | 0.0900 |
| X16 | -0.3207 | 0.1772 | 0.0712 | -0.2061 | -0.1148 | -0.0331 | -0.0384 |
| X17 | -0.3672 | -0.0052 | -0.1237 | -0.1118 | 0.0200 | 0.0910 | -0.0173 |
| X18 | -0.3807 | -0.0116 | 0.0603 | -0.1271 | -0.0378 | 0.0219 | -0.1220 |
| X19 | -0.3523 | 0.0702 | 0.0201 | -0.2240 | 0.1222 | -0.0485 | -0.0747 |

3.3.2. 交叉验证

由于本数据集变量众多, 为了获得更好地预测分类效果(当在所有变量存在和降维之后进行预测分类, 逻辑回归都无法达到收敛), 本文利用交叉验证进行了子集选择。根据预测误差最小原则, 选取 X11、X18 两个特征作为预测变量。

4. 实证分析

4.1. 逻辑回归(LR)

由于使用降维后的数据作逻辑回归,算法无法收敛,因此本文只呈现交叉验证后的数据逻辑回归的结果。

Table 6. LR confusion matrix (cross validation)

表 6. LR 混淆矩阵(交叉验证)

| 预测结果 | 实际结果 | |
|------|------|-----|
| | 患病 | 不患病 |
| 患病 | 40 | 13 |
| 不患病 | 11 | 8 |

根据混淆矩阵表 6, 计算逻辑回归的分类准确率, 准确率 = $(40 + 8)/(40 + 13 + 11 + 8) = 66.67\%$, 分类准确率一般。

4.2. 线性判别分析

Table 7. LDA confusion matrix (cross validation)

表 7. LDA 混淆矩阵(交叉验证)

| 预测结果 | 实际结果 | |
|------|------|-----|
| | 患病 | 不患病 |
| 患病 | 23 | 2 |
| 不患病 | 2 | 8 |

Table 8. LDA confusion matrix (dimension reduction)

表 8. LDA 混淆矩阵(降维)

| 预测结果 | 实际结果 | |
|------|------|-----|
| | 患病 | 不患病 |
| 患病 | 21 | 3 |
| 不患病 | 4 | 7 |

根据表 7 和表 8 可以分别计算交叉验证子集选择后数据预测分类准确率 = $(23 + 8)/(23 + 2 + 2 + 8) = 88.57\%$; 降维后数据预测分类准确率 = $(21 + 7)/(21 + 3 + 4 + 7) = 80.00\%$ 。显然, 交叉验证子集选择后数据预测分类效果要好于降维后数据的预测分类效果。

4.3. 二次判别分析(QDA)

Table 9. QDA confusion matrix (cross validation)

表 9. QDA 混淆矩阵(交叉验证)

| 预测结果 | 实际结果 | |
|------|------|-----|
| | 患病 | 不患病 |
| 患病 | 24 | 2 |
| 不患病 | 1 | 8 |

Table 10. QDA confusion matrix (dimension reduction)**表 10.** QDA 混淆矩阵(降维)

| 预测结果 | 实际结果 | |
|------|------|-----|
| | 患病 | 不患病 |
| 患病 | 23 | 3 |
| 不患病 | 2 | 7 |

根据表 9 和表 10 可以分别计算交叉验证子集选择后数据预测分类准确率 $= (24 + 8)/(24 + 2 + 1 + 8) = 91.43\%$ ；降维后数据预测分类准确率 $= (23 + 7)/(23 + 3 + 2 + 7) = 85.71\%$ 。

利用二次判别分析进行预测分类，两组数据的准确率较线性判别分析有明显上升，特别是交叉验证组，其预测分类的准确率已超过 90%，表明数据变量可能是来自不同方差的总体。

4.4. *K*-近邻(*K*-NN)

Table 11. *K*-NN confusion matrix (cross validation)**表 11.** *K*-NN 混淆矩阵(交叉验证)

| | 预测结果 | 实际结果 | |
|--------------|------|------|-----|
| | | 患病 | 不患病 |
| <i>K</i> = 1 | 患病 | 24 | 3 |
| | 不患病 | 1 | 7 |
| <i>K</i> = 2 | 患病 | 23 | 3 |
| | 不患病 | 2 | 7 |
| <i>K</i> = 3 | 患病 | 23 | 3 |
| | 不患病 | 2 | 7 |

Table 12. *K*-NN confusion matrix (dimension reduction)**表 12.** *K*-NN 混淆矩阵(降维)

| | 预测结果 | 实际结果 | |
|--------------|------|------|-----|
| | | 患病 | 不患病 |
| <i>K</i> = 1 | 患病 | 22 | 2 |
| | 不患病 | 3 | 8 |
| <i>K</i> = 2 | 患病 | 22 | 1 |
| | 不患病 | 3 | 9 |
| <i>K</i> = 3 | 患病 | 23 | 2 |
| | 不患病 | 2 | 8 |

当 *K* 取 1,2,3 时，根据表 11 和表 12 可以计算出交叉验证子集选择后数据预测分类准确率分别为 88.57%、85.71%、85.71%，后面两种预测分类结果完全一致；降维后数据预测分类准确率分别为 85.71%、88.57%、88.57%，两组数据分类结果差距不大。

4.5. 朴素贝叶斯(NB)

Table 13. NB confusion matrix (cross validation)

表 13. NB 混淆矩阵(交叉验证)

| 预测结果 | 实际结果 | |
|------|------|-----|
| | 患病 | 不患病 |
| 患病 | 24 | 2 |
| 不患病 | 1 | 8 |

Table 14. NB confusion matrix (dimension reduction)

表 14. NB 混淆矩阵(降维)

| 预测结果 | 实际结果 | |
|------|------|-----|
| | 患病 | 不患病 |
| 患病 | 24 | 3 |
| 不患病 | 1 | 7 |

根据表 13 和表 14 分别计算出交叉验证子集选择后数据预测分类准确率 = $(24 + 8)/(24 + 2 + 1 + 8) = 91.43\%$; 降维后数据预测分类准确率 = $(24 + 7)/(24 + 3 + 1 + 7) = 88.57\%$ 。

在本次实证分析中, 朴素贝叶斯分类器的准确率最高, 同时表明所研究的变量可能来自高斯分布。

4.6. 实验结果总结

通过 5 种分类器进行判别分析, 可以发现:

- 1) 当特征变量比较大时, 逻辑回归算法很难达到收敛, 此时要进行降维和变量选择;
- 2) 在预测效果方面, 贝叶斯判别效果最佳, 其次是二次判别分析;
- 3) 针对不同的数据集, 几种判别分析表现可能存在差异; 因此在进行预测分类前, 需要对数据内部进行分析, 了解它们之间的相关性和近似分布。

5. 结论

宫颈癌作为全世界女性的严重公共卫生问题, 幸运的是人们可以通过技术手段进行预防。而通过行为预防对宫颈癌进行一级预防具有许多优势[9], 如果及早发现, 可以降低死亡率并提高治疗效果。

根据实验结果, 本文的模型具有一定的预测能力, 且部分模型的预测分类准确率已达到 90%以上, 表明机器学习在疾病检测方面有充分的实践可能性。同时也表明了行为决定因素成为参与宫颈癌一级预防的重要方面, 且这种预防方式更彻底、更具经济性[10]。

参考文献

- [1] 冷雪娇, 吴沁航, 王卓. 宫颈癌预防及治疗研究进展[J]. 现代医药卫生, 2021, 37(24): 4241-4245.
- [2] 徐宁志. 人类宫颈癌防治的挑战和愿景[J]. 张江科技评论, 2021(6): 20-23.
- [3] 曹蕾, 杨丽云, 程欣, 朱佩茹. 育龄期女性宫颈癌筛查前焦虑状况及其影响因素分析[J]. 护理实践与研究, 2021, 18(24): 3645-3649.
- [4] 周琦, 吴小华, 刘继红, 李力, 朱笕青, 白萍, 盛修贵. 宫颈癌诊断与治疗指南(第四版) [J]. 中国实用妇科与产科杂志, 2018, 34(6): 613-622.
- [5] 刘萍. 中国大陆 13 年宫颈癌临床流行病学大数据评价[J]. 中国实用妇科与产科杂志, 2018, 34(1): 41-45.

-
- [6] 乔友林, 赵宇倩. 宫颈癌的流行病学现状和预防[J]. 中华妇幼临床医学杂志(电子版), 2015, 11(2): 1-6.
- [7] 张燕茹, 王月云, 刘植华. 宫颈癌防治研究进展[J]. 中国肿瘤, 2015, 24(12): 998-1002.
- [8] 李荣. 宫颈癌根治术后患者的病耻感及其与生活质量的相关性[D]: [硕士学位论文]. 济南: 山东大学, 2015.
- [9] Sobar, Machmud, R. and Wijaya, A. (2016) Behavior Determinant Based Cervical Cancer Early Detection with Machine Learning Algorithm. *Journal of Computational and Theoretical Nanoscience*, **22**, 3120-3123.
<https://doi.org/10.1166/asl.2016.7980>
- [10] 程静新, 姚立丽, 李贺月, 袁敏, 周萍, 袁建林, 张怡. 5766 例宫颈癌临床特征分析[J]. 实用妇产科杂志, 2014, 30(10): 768-772.