

一种基于局部敏感哈希的SVM快速增量学习算法

姚明海 林宣民 王宪保

(浙江工业大学信息工程学院 杭州 310032)

摘要 为了提高大规模高维度数据的训练速度和分类精度,提出了一种基于局部敏感哈希的SVM快速增量学习方法。算法首先利用局部敏感哈希能快速查找相似数据的特性,在SVM算法的基础上筛选出增量中可能成为SV的样本,然后将这些样本与已有SV一起作为后续训练的基础。使用多个数据集对该算法进行了验证。实验表明,在大规模增量数据样本中,提出的SVM快速增量学习算法能有效地提高训练学习的速度,并能保持有效的准确率。

关键词 LSH,SVM,增量学习,大规模数据,高维

中图分类号 TP181 文献标识码 A

Fast Incremental Learning Algorithm of SVM with Locality Sensitive Hashing

YAO Ming-hai LIN Xuan-min WANG Xian-bao

(College of Information Engineering, Zhejiang University of Technology, Hangzhou 310032, China)

Abstract In order to improve the training speed and the classification accuracy in large scale high dimension data, a new incremental learning algorithm of SVM with LSH was proposed. It uses the LSH algorithm, which can seek similar data fast in a large scale and high dimension data, to filter out the incremental samples which may become SVs on the basis of the SVM algorithm. Then it makes the selected samples and the existing SVs as a basis for the following training. We took advantages of the multiple data sets to validate the algorithm. Experiments show that this new algorithm can improve the speed of the incremental training learning in large scale data with the effective accuracy.

Keywords LSH,SVM,Incremental learning, Large scale data, High dimension

1 引言

近年来,随着计算机网络技术和传感器技术的发展,应用软件和设备产生了大量的高维数据,且规模呈几何级数增加。若要从这些数据中学习、抽取有用的知识,至少有两个主要问题需要解决:1)如何处理不断增长的数据,同时保持学习训练得到模型的高性能;2)如何快速适应动态环境。第一个问题可以通过增量学习的重要信息来解决^[1];对于第二个问题,可以通过快速删除在学习中不重要的数据来减少学习时间。

支持向量机^[2](Support Vector Machine, SVM)是一种以统计学习理论为基础的机器学习技术,主要应用于求解监督学习问题。在实际应用中,训练样本常常表现出上述特点,这就要求分类器不断地对新样本重新进行训练。而增量学习旨在获得原训练样本与新增样本并集的最优解。因此,对于不断增长的数据集而言,SVM增量学习具有独特的优势。其中,固定划分增量学习算法(Batch Incremental SVM, Batch-ISVM)是最早的基于SVM的增量学习算法,其基本思想是:每次增量过程只保留支持向量集(Support Vectors, SV),舍弃非支持向量集,并将本次SV集与所有新增样本作为下一次增量学习的训练样本。但事实上,这样没有对新增样本进行筛选,将其中没用或不重要的样本也进行了训练,降低了增

量训练的速度;同时,其中的一些糟糕的样本也可能会影响精度。文献[4]引入卡罗需-库恩-塔克条件(Karush-Kuhn-Tucker, KKT)条件,来提取新增样本中的有用信息,即基于卡罗需-库恩-塔克条件的SVM增量学习(Incremental SVM with Karush-Kuhn-Tucker, KKT-ISVM)。后来的文献[5-7]都是基于KKT条件进行改进。文献[8]引入了错误驱动的概念,把错分类样本也考虑在内。而文献[9]提出新的基于KKT条件的错误驱动SVM增量学习。这些增量学习大部分利用KKT条件来筛选新增样本,筛选速度慢,影响了训练的总速度。当然,近几年来也有很多研究将SVM增量学习应用于特定领域对象。例如,Jhala等^[10]将SVM增量学习用于识别垃圾邮件;文献[11]则将自联想神经网络与SVM增量结合用于视频事件检测;Rupesh等^[12]则将其用于人体活动的识别。

上述SVM增量学习方法虽然都能对大规模高维数据进行学习,但都存在不足之处:1)不筛选新增样本,直接与上次SV集一起训练,增加了训练时间;2)筛选时间太长,影响训练的总时间;3)应用范围小。本文将局部敏感哈希(Locality Sensitive Hashing, LSH)思想方法与SVM增量学习思想相结合,提出一种基于主成分分析的局部敏感哈希SVM快速增量学习算法(Incremental SVM with Principal Component

本文受面向非特定产品质量检测的一般性目标识别方法(LZ14F030001)资助。

姚明海(1963-),教授,博士生导师,主要研究方向为机器学习与缺陷检测, E-mail: ymh@zjut.edu.cn; 林宣民(1991-),男,硕士生,主要研究方向为机器学习与缺陷检测, E-mail: x790144106@163.com; 王宪保(1977-),男,博士,讲师,主要研究方向为机器学习与深度学习, E-mail: wxb@zjut.edu.cn.

Analysis of Locality Sensitive Hashing, PCA-LSH-ISVM)。此算法能快速筛选出大规模高维数据的重要样本,同时改善训练速度,以缩短总体训练时间。实验结果表明,本文算法在加快学习训练大规模高维样本的速度的同时,具有较好的分类效果,是一种有效的增量 SVM 学习算法。

2 SVM 和 LSH 理论基础

2.1 SVM 理论基础和 KKT 条件

SVM 的核心思想是在样本空间中找到一个分类超平面,使得分类间隔最大化。若给定样本集 $\{(x_i, y_i)\}, i=1, 2, \dots, n, x_i \in R^m, y_i \in \{-1, +1\}$, 则求最优分类面可归结为求解如下二次规划问题:

$$\begin{aligned} \max Q(\alpha) &= \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j x_i^T x_j \\ \text{s. t. } & \sum_{i=1}^n \alpha_i y_i = 0, 0 \leq \alpha_i \leq C, i=1, 2, \dots, n \end{aligned} \quad (1)$$

其中, α_i, α_j 为拉格朗日乘子; C 为惩罚因子; y_i, y_j 为样本的类别标签; 最优解 $\alpha = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_n]$, 使得每一个样本满足 KKT 条件:

$$\begin{cases} \alpha_i = 0 \Rightarrow y_i, & f(x_i) \geq 1 \\ 0 < \alpha_i < C \Rightarrow y_i, & f(x_i) = 1 \\ \alpha_i = C \Rightarrow y_i, & f(x_i) \leq 1 \end{cases} \quad (2)$$

其中, $f(x_i)$ 为决策函数。非 0 的 α_i 所对应的样本组成 SV。

SVM 具有一个独特的特点,它使用训练实例的一个子集来表示决策边界,即 SV 集。SV 集可以充分描述整个训练样本集的特征,与训练样本集是等价的^[4]。然而,新增样本的加入,可能打破样本集与 SV 集的等价关系,这就需要重新寻找 SV 集,并求出新的最优超平面。

由求解最优化问题的过程可知,处在类别边缘地带并且距离其他类别较近的样本成为新的 SV 的可能性最大。如图 1 所示,原分类面为 $f(x)=0$, 支持向量为 s_1, s_2 。当有新增样本 $\{p_1, p_2, n_1, n_2\}$ 加入后, $f'(x)=0$ 成为新的分类面, $\{s_1, s_2, a, b, c, d, p_2\}$ 为新的支持向量集。显然,原来的 $\{a, b, c, d\}$ 4 个非 SV 样本转化为新的 SV, 而它们正是处于类别边缘地带并且距分类面和其他类别较近的样本。

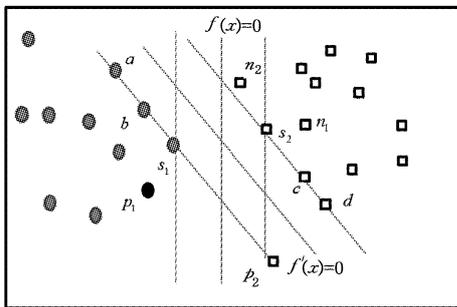


图 1 新 SV 转化示意图

SVM 增量学习就是要充分利用历史训练结果,在学习过程中选择最有可能成为新 SV 的样本,舍弃对分类结果没有贡献的历史样本,从而达到加快训练速度、减少存储空间的目的。

2.2 LSH 理论基础

LSH 算法的基本思想是:两个在原始空间中的邻近或相似的数据样本点进行相同的映射或投影变换,得到的数据点在新空间中仍然相邻或相似的概率很大,同时不相邻的数据

点被映射到新空间中相似的概率很小。利用一组哈希函数^[13]对原空间集合中的数据进行哈希投影后,就得到了一个哈希桶(即哈希表)。这样,原数据集合被分成了多个哈希桶,同时这些哈希桶中的数据是相邻或相似的,且该哈希桶的数据样本个数较少。因此,在超大数据集中搜索相邻或相似的数据样本的问题,就变成了在哈希桶中的少量数据样本中搜索的问题。

哈希函数需要满足以下两个条件:

- (1) 如果 $d(x, y) \leq d_1$, 则 $h(x) = h(y)$ 的概率至少为 p_1 ;
- (2) 如果 $d(x, y) \geq d_2$, 则 $h(x) = h(y)$ 的概率至多为 p_2 。

其中, $d(x, y)$ 为 x 和 y 之间的距离; d_1, d_2 为距离阈值; $h(x)$ 和 $h(y)$ 为表示 x 和 y 的哈希表。

以上两个条件都满足的哈希函数称为 (d_1, d_2, p_1, p_2) 哈希敏感函数。通过一个或多个 (d_1, d_2, p_1, p_2) 敏感的哈希函数对原始数据集合进行哈希变换,从而生成一个或多个哈希表的过程,称为局部敏感哈希。局部敏感哈希函数族有多种,如汉明距离下的最小哈希函数族^[14]、欧氏距离下的基于 p -稳定分布的哈希函数族^[15]等,它们是根据不同距离量度进行划分的。

3 基于 LSH 的 SVM 快速增量学习

增量学习的主要任务是利用历史结果避免重复训练,以得到较高精度的分类结果。传统的 SVM 在增量学习时主要是利用已有 SV 集,然后与新增样本一起训练。基于 KKT 条件的 SVM 增量学习则利用得到的 SV 和经 KKT 条件筛选的新增样本一起训练。这两种方法对于大规模高维新增样本的训练速度较慢。LSH 能通过上次训练得到的 SV 集,快速筛选出新增样本中与 SV 邻近的样本,它们既包含不满足 KKT 条件的样本,也包含满足 KKT 条件但邻近 SV 的样本。因此,本文基于 LSH 的 SVM 增量学习算法既保证了分类器的精度,又使训练速度得到提高。

3.1 PCA-LSH 的构建

本文利用主成分分析 (Principal Component Analysis, PCA) 投影得到的特征向量进行 hash 编码投影,而不是随机的向量。这种特征子空间提供了很好的近似输入空间,逼近误差可以通过累积输入特征成分的近似子空间占整个输入空间的比率来控制。因此,通过 PCA 来替代 LSH 投影的随机向量,可以很容易从一个给定的数据分布来确定一个合适的 hash 函数数量。

设 PCA 所得到的特征投影向量的数量为 N , 即特征投影向量表示为 $U_N = \{u_1, u_2, \dots, u_N\}$, 投影得到的向量表示为 $V = \{v_1, v_2, \dots, v_N\}$, x 为输入样本向量, \bar{x} 为输入样本向量的均值。则:

$$V = U_N^T (x - \bar{x}) \quad (3)$$

设 hash 编码的一维空间向量区域为 $[1, P]$, $F(v_i)$ 为 hash 函数, 则 $1 \leq F(v_i) \leq P$ 。

$$F(v_i) = \max\{\text{round}\left(\frac{\min(\max(v_i, v_{i-\min}), v_{i-\max}) - v_{i-\min}}{v_{i-\max} - v_{i-\min}} P\right), 1\} \quad (4)$$

其中, $v_{i-\min}, v_{i-\max}$ 表示初始输入样本投影后特征向量的第 i 个特征值的最小值和最大值; $\text{round}()$ 为四舍五入函数。

因此,输入特征向量的 hash 值向量由如下 $h(x)$ 表示:

$$h(x) = \{F(v_1), F(v_2), \dots, F(v_i)\} \quad (5)$$

然后,利用 $h(x)$ 组成 hash 表 H :

$$H = [h(x_1), h(x_2), \dots, h(x_n)] \quad (6)$$

最后,利用 $h(x)$, H 对新增样本进行筛选。

3.2 基于 LSH 的 ISVM 学习算法

本文提出了一种基于 PCA-LSH 的 SVM 增量学习方法 PCA-LSH-ISVM,算法利用已有 SV 集和 PCA-LSH 方法筛选增量样本,然后再把得到的样本和 SV 集一起训练。结果表明,在保持精度的前提下,该算法有效提升了训练速度。

假设初始样本集为 Ω_0 ,新增样本集为 Ω_1 ,初始样本训练得到的 SV 集为 Ω_{0SV} ,增量学习后得到的 SV 集为 Ω_{1SV} ,初始样本训练得到的 SV 集 LSH 后的 hash 表为 H_0 ,增量学习后得到的 SV 集 LSH 后的 hash 表为 H_1 ,更新后的 hash 表为 H ,新增样本 LSH 后的 hash 值向量为 $h(\Omega_1)$,新增样本筛选后的样本为 Ω_{1LSH} 。则 PCA-LSH-ISVM 算法的详细描述如下:

(1)用 Ω_0 训练得到 SV 集 Ω_{0SV} 。

(2)基于得到的 Ω_{0SV} ,通过 PCA-LSH 算法建立 hash 表 H_0 。

(3)对于 Ω_1 ,通过 PCA-LSH 算法 hash 函数得到每个样本对应的 hash 值向量 $h(\Omega_1)$ 。

(4)筛选新增样本集 Ω_1 :如果 Ω_1 中样本对应的 $h(\Omega_1)$ 落在 H_0 中,则得到筛选后的样本集 Ω_{1LSH} ;如果不存在 Ω_{1LSH} ,则直接跳到下一个新增样本集进行筛选。

(5)将 Ω_{0SV} 和 Ω_{1LSH} 作为下一次 SVM 学习的训练集,训练得到 Ω_{1SV} 。

(6)更新 hash 表:将 Ω_{1SV} 通过 PCA-LSH 得到的 hash 表 H_1 更新 H_0 ,生成 H 。

(7)将 H 对下一个新增样本集得到筛选后的样本与上次训练得到的 SV 一起学习训练,更新 hash 表,直到没有新增样本集输入。

4 实验

为了验证算法的有效性,将 PCA-LSH-ISVM 与 SVM, Batch-ISVM^[4], KKT-ISVM^[7] 算法作对比。实验在 UCI 标准数据集 Steel、Waveform、Magic、default of credit card clients (Credit)、CASp、人脸识别数据集 PIE Face Database of CMU (PIEFace)、手写数字数据集 MNIST (Mnist) 和频繁模式挖掘评测常用数据集 Pumsb 上分别进行。因为算法是针对大规模数据,一般为非线性,所以采用核函数 (RBF) 的 SVM。实验将样本数据集随机分成初始样本集,增量样本集 1—6 和测试集。数据集个数、特征个数和增量样本的详情如表 1 所列。实验在 i5 M430 处理器进行,主频为 2.27GHz,内存为 4GB,软件为 PyCharm5.0.3。

表 1 数据集和增量参数

数据集	总数	特征个数	初始样本	增量样本	测试样本
Steel	1075	26	200	100	275
Waveform	5000	21	500	500	1500
Magic	19020	10	1000	2000	6020
Credit	30000	23	1000	3000	11000
Casp	45730	9	2000	5000	13730
PIEFace	5780	1024	550	550	1930
Mnist	4366	784	400	400	1566
Pumsb	32476	73	3000	3300	9676

4.1 实验结果及分析

在哈希函数参数 $P=3$ 的情况下,对比各算法的增量训

练速度和分类精度。其中,本文 PCA-LSH 只是对新增样本进行筛选,并不改变样本原来的属性(未进行其他预处理),这样就保证了与其他算法训练学习对象样本的一致性。实验结果如表 2、表 3 所列。

表 2 不同算法的分类精度/%

数据集	SVM	Simple-ISVM	KKT-ISVM	PCA-LSH-ISVM
Steel	78.91	78.55	78.10	78.91
Waveform	88.27	88.53	88.47	88.60
Magic	86.68	86.59	86.69	86.77
Credit	81.73	81.65	81.66	81.75
Casp	76.40	76.53	76.35	76.56
PIEFace	88.76	88.76	88.76	89.27
Mnist	83.46	83.46	83.46	83.46
Pumsb	96.28	96.26	96.29	96.29

表 3 不同算法的增量训练时间/s

数据集	SVM	Simple-ISVM	KKT-ISVM	PCA-LSH-ISVM
Steel	0.250	0.162	0.399	0.160
Waveform	1.651	0.706	1.051	0.686
Magic	16.11	7.61	9.25	5.63
Credit	107.51	40.76	45.05	26.98
Casp	140.33	90.50	92.32	63.25
PIEFace	93.04	91.58	104.78	80.16
Mnist	50.48	44.54	38.48	27.08
Pumsb	394.52	64.12	63.77	44.73

表 3 中的增量训练时间是指筛选新增样本的时间和进行学习训练的时间之和。其中,SVM, Batch-ISVM 未对新增样本进行筛选,所以没有筛选时间,只有训练时间。

由上述实验结果可知:

1)本文算法在保持较高分类精度的同时,增量训练的速度具有较大优势,相对于其他最快算法平均提升了约 20%;

2)在分类精度上,SVM 和 Batch-ISVM 都没有对新增样本进行筛选,其中的一些不好的样本在一定程度上影响了分类的精度,如在 magic 数据集上两者的精度分别为 86.68% 和 86.59%,都低于 KKT 筛选算法的 86.69% 和本文算法的 86.77%;

3)KKT-ISVM 只是筛选出违反 KKT 条件的新增样本,将新增样本中可能成为 SV 但满足 KKT 条件的样本删除了,以至于影响了分类精度;

4)本文算法筛选与上次 SV 相似或相邻的样本(包含违反和满足 KKT 条件但可能成为新的 SV 的样本),因而分类精度有一定程度的提高。

5)与其他算法相比,在新增样本数量维度较小时,其训练速度提高不明显,相对于其他最快算法只平均提高了 2% 左右;在增量样本数量或维度较大时,其训练的速度就有明显优势,相对于其他最快算法平均提高了 27% 左右。

6)由表 2、表 3 可知,在 UCI 标准数据集、人脸识别数据集、手写数字数据集和频繁模式挖掘评测常用数据集中,本文算法都适用。因此,本文算法在不同数据集中有一定通用性。

对于 SVM 而言,其不对新增样本和初始样本进行筛选,相当于新增样本进来一次就重新训练一次,这大大增加了训练时间;Batch-ISVM 让上次训练得到的 SV 与新增样本直接训练,导致了增量训练时间也较长;KKT-ISVM 虽然对新增样本进行是否满足 KKT 条件的筛选,但它在大规模高维数据样本中的筛选速度太慢,以至于总体的增量训练时间过长;本文算法则能对大规模高维数据进行快速的搜索筛选,提高

了总体增量学习的训练速度。

结束语 本文基于局部敏感哈希在大规模高维数据中快速搜索的特性,提出了一种基于主成分分析的局部敏感哈希的 SVM 快速增量学习算法 PCA-LSH-ISVM。该算法将主成分分析应用于局部敏感哈希得到特定数量的哈希函数和上次训练得到的 SV 集样本生成哈希表,然后对新增样本进行筛选,再将筛选得到的新增样本和上次训练得到的 SV 集样本作为下次的训练样本进行训练学习。实验结果表明,在大规模高维增量学习样本中,该算法不仅能保证分类精度和良好的推广能力,而且学习训练速度也比经典的 SVM 增量学习算法快,可以快速进行增量学习。

参 考 文 献

- [1] ATTAR V, SINHA P, WANKHADE K. A fast and light classifier for data streams[J]. *Evol Syst*, 2010, 1(3): 199-207.
- [2] VAPNIK V. *The Nature of Statistical Learning Theory*[M]. New York: Springer Verlag, 1995.
- [3] SYED N A, SUNG K. Handling Concept Drifts in Incremental Learning with Support Vector Machines[C]//Proc. of the 5th ACM SIGKDD International Conference. 1999: 316-321.
- [4] DIEHL C P, CAUWENBERGHS G. SVM Incremental Learning Adaptation and Optimization [C] // International Joint Conference on Neural Networks. IEEE, 2003: 2685-2691.
- [5] 茅嫣蕾, 魏赟. 一种基于 KKT 条件和壳向量的 SVM 增量学习算法[J]. *电子科技*, 2016, 29(2): 38-40.
- [6] 李妍坊, 苏波, 刘功申. 一种基于组合保留集的 SVM 增量学习算法[J]. *上海交通大学学报*, 2016, 50(7): 1054-1059.
- [7] 曹健, 孙世宇, 段修生, 等. 基于 KKT 条件的 SVM 增量学习

算法[J]. *火力与指挥控制*, 2014(7): 139-143.

- [8] LUO J, PRONOBIS A, CAPUTO B. Incremental Learning for Place Recognition in Dynamic Environments[C]//IEEE International Conference on Intelligent Robots and Systems. 2007: 721-729.
- [9] 张灿淋, 姚明海, 童小龙, 等. 一种新的基于 KKT 条件的错误驱动 SVM 增量学习算法[J]. *计算机系统与应用*, 2014, 23(1): 144-148.
- [10] JHALA I S, DALAL P. Optimized Incremental SVM based Classifier for Spam Filtering using Internet Acronyms[J]. *International Journal for Innovative Research in Science & Technology*, 2015, 2(1): 2349-6010.
- [11] CHAKROUN M, WALI A, ARIBI Y, et al. Video event detection using auto-associative neural network and incremental SVM models[C]//International Conference on Intelligent System Design & Application. 2015: 563-568.
- [12] JAGTAP R V, POTEY M A. Recognition of Human Activity using Incremental SVM[J]. *Imperial Journal of Interdisciplinary Research*, 2016, 7(2): 2454-2462.
- [13] ANDONI A, INDYK P, NGUYEN H L. Beyond locality-sensitive hash [C]//Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms. 2014: 1018-1028.
- [14] GIONIS A, INDYK P, MOTWANI R. Similarity search in high dimensions via hashing[C]//Proceedings of the International Conference on Very Large Databases. 1999: 25-38.
- [15] DATAR M, IMMORLICA N, INDYK P, et al. Locality sensitive hashing scheme based on p-stable distributions [C]//Proceedings of the ACM Symposium on Computational Geometry. 2004: 23-36.

(上接第 67 页)

对比图 2—图 4 可以得到以下结论: 1) 降维后的预测精度明显高于降维前, 计算图 2 所示精度平均值, 结果显示平均提高了 15.07%; 2) 随着隐藏节点数的增加, 降维前和降维后的预测精度都有所提高, 但降维后的提高速度要优于前者; 3) BP 神经网络在降维后收敛的速度快于降维前。

结束语 慢性肝病的成因复杂, 医学检查指标多, 指标之间存在信息冗余和干扰, 导致准确地智能预测慢性肝病非常困难。为了提高智能预测慢性肝病的准确度并降低系统的复杂性, 本文利用 ROC 特征曲线优选出 13 维敏感度较高的指标, 并利用主成分分析方法基于优选指标将维度进一步降至 5 维, 最后利用 BP 神经网络构建出智能预测慢性肝病模型。从 125 组数据中随机选出 115 组作为训练样本集合进行 BP 神经网络的训练, 并将剩余的 10 组数据作为测试样本, 用于验证智能预测模型的准确度。本文大量地重复上述训练和预测过程, 以平均值作为最终的准确度, 以提高智能预测模型的可信度。在同等条件下, 对未经过降维处理的原始数据进行 BP 神经网络训练和预测, 实验表明前者不但降低了系统的复杂性, 而且提高了 15.07% 的准确度。因此, 针对慢性肝病, 本文的智能预测模型能达到一定的精度和准确度, 为医生的辅助诊断提供可行的方法。

参 考 文 献

- [1] 世界卫生组织[OL]. <http://www.who.int/campaigns/hepatitis-day/2016/event/zh>.
- [2] 王恩成, 唐琳, 王健, 等. 慢性乙型肝炎中医症候聚类分析研究[J]. *中国中西医结合杂志*, 2014, 34(1): 39-42.
- [3] HO C Y, LAI Y C, CHEN I W, et al. Statistical Analysis of False Positives and False Negatives from Real Traffic with Intrusion Detection/Prevention Systems[J]. *Communications Magazine*, IEEE, 2012, 50(3): 146-154.
- [4] PAXSON V, ASANOVIC K, DHARMAPURIKAR S, et al. Rethinking hardware support for network analysis and intrusion prevention[C]//Proc of the 15th USENIX Workshop on Hot Topics in Security. Berkeley, CA: USENIX, 2006: 63-68.
- [5] THOMPSON K. Programming techniques: Regular expression search algorithm[J]. *Communications of the ACM*, 1968, 11(6): 419-422.
- [6] RABIN M O, SCOTT D. Finite automata and their decision problems[J]. *IBM Journal of Research & Development*, 1959, 3(2): 114-125.
- [7] 高秀娟. BP 神经网络在肝硬化治疗预测中的应用[J]. *数学理论与应用*, 2011, 31(3): 21-23.
- [8] 李耐萍. 基于机器算法的肝纤维化无创诊断分析与比较[D]. 长沙: 中南大学, 2013: 30-36.