0254-6124/2022/42(3)-0340-06 *Chin. J. Space Sci.* 空间科学学报

GONG Zhe, ZOU Ziming, LU Yang. Solar Proton Events Short-time Forecasting Based on Ensemble Learning (in Chinese). *Chinese Journal of Space Science*, 2022, **42**(3): 340–345. DOI:10.11728/cjss2022.03.210310025

基于集成学习的太阳质子事件短期预报方法*

官 哲[1,2] 邹自明[1] 陆 阳[1]

1(中国科学院国家空间科学中心 北京 100190)

2(中国科学院大学 北京 100049)

摘 要 太阳质子事件是一种由太阳活动爆发时喷射并传播到近地空间的高能粒子引起的空间天气现象。这些高能粒子会对航天器和宇航员产生严重危害，对太阳质子事件进行准确的短期预报是航天活动灾害预防的重要内容。针对当前主要预报模型中普遍存在的高虚报率问题，提出了一种基于集成学习的太阳质子事件短期预报方法，利用第 23 个太阳活动周数据，建立了一种集成 8 种机器学习模型的太阳质子事件短期预报系统。实验结果表明，本文方法在取得了 80.95% 的报准率的同时，将虚报率降低至 19.05%，相比现有的预报系统具有较为明显的优势。

关键词 太阳质子事件，短期预报，集成学习，虚报率

中图分类号 P353

Solar Proton Events Short-time Forecasting Based on Ensemble Learning

GONG Zhe[1,2] ZOU Ziming[1] LU Yang[1]

1(*National Space Science Center*, *Chinese Academy of Sciences*, *Beijing* 100190)

2(*University of Chinese Academy of Sciences*, *Beijing* 100049)

Abstract Solar proton event is a space weather phenomenon caused by energetic particles ejected and propagated into near-Earth space during bursts of solar activity. These high-energy particles can cause serious harm to spacecraft and astronauts, therefore, accurate short-term forecasting of solar proton events is very necessary as part of disaster prevention for space activities. The short-time forecasting of solar proton events still faces a lot of challenges, one of which is the high false alarm rate. To solve this problem, we adopted a whole new set of methods-machine learning. As a branch of computer science, many excellent algorithms have emerged in the field of machine learning in recent years, and have achieved successful applications in many fields. In this study, an ensemble model based on 8 widely used machine learning models is established to make precise forecasting of solar proton events. An experiment

* 中国科学院“十三五”信息化建设专项资助（XXH13505-04）

2021-03-09 收到原稿，2022-04-01 收到修定稿

E-mail：PhilGong@126.com

on the 23rd solar cycle shows that this model gets a probability of detection of 80.95% and a false alarm rate of 19.05%.

Key words　Solar proton events, Short-time prediction, Ensemble learning, False alarm rate

0　引言

太阳质子事件是一种空间天气现象。太阳活动例如耀斑和日冕物质抛射(CME)等爆发时,会喷射出大量的高能粒子,其中一部分会从太阳低层大气传播到近地空间,并被地球附近的卫星或飞行器探测到[1]。当高能粒子通量达到一定标准时,便认为发生了太阳质子事件。中国空间环境预测中心将太阳质子事件定义为:在地球同步轨道上探测能量大于10 MeV的质子,每5 min计算一次平均积分通量,单位pfu(1 pfu = 1 $cm^{-2}{\cdot}s^{-1}{\cdot}sr^{-1}$)。当连续三次积分通量大于10 pfu时,便认为发生了太阳质子事件。太阳质子事件中,大量的高能粒子会对航天器和宇航员产生严重危害,因此准确的太阳质子事件短期预报对航天活动灾害预防是非常必要的。

根据预报方法的不同,可以将太阳质子事件的预报模式分为物理预报模式和经验预报模式两种。物理预报模式是基于太阳质子事件发生的物理机制建立预报模型。例如,SOLPENCO模型[1]根据从太阳质子事件开始到激波到达拉格朗日点(L1)的时间计算质子事件的通量变化趋势对太阳质子事件进行预报;SEPMOD模型[2]通过计算给定位置的高能粒子通量曲线实现太阳质子事件的预报。经验预报模式是基于质子事件相关的太阳活动与现象,例如利用太阳黑子,太阳耀斑,日冕物质抛射等,对太阳质子事件进行预报。2006年美国空军实验室(AFRL)以IMP系列卫星于1997－2001年观测到的21个质子事件与对应耀斑之间的关系为基础,建立了质子预报系统(Proton Prediction System, PPS)[3],其报准率能达到85.71%,但是虚报率高达50%;2009年,Laurenza等[4]选取了1995－2005年的70个太阳质子事件,以耀斑经纬度、耀斑重要性等级、软X射线积分通量和射电噪暴积分通量等作为输入建立预报模型,该预报模型将虚报率降低至42%的同时,报准率也下降至63%;Núñez [5]提出了UMASEP模型,该模型通过分析第22和第23太阳活动周期间软X射线和近地空间的9～500 MeV的微分质子通量等数据,预报太阳质子事件发生与否,其报准率为80.72%,虚报率为33.99%。Huang等[6]使用集成学习的方法,基于耀斑和CME相关数据建立了两个隐朴素贝叶斯模型,并将这两个模型集成对太阳质子事件进行预测,取得了78.57%的报准率和50%的虚报率。Zhong等[7]通过分析地球静止运行环境卫星5 min平均软X射线通量和差分积分质子通量,预测在未来24 h内是否会发生太阳质子事件,其报准率为80.00%,虚报率为25.85%。

现存太阳质子事件短期预报方法的局限性主要在于:在获得较高的报准率时,虚报率依然较高。针对这个问题,本文提出一种基于集成学习的太阳质子事件短期预报方法,选取了8种具有不同特点的机器学习模型,以现存方法中广泛使用的太阳黑子数据、耀斑数据、CME数据作为输入对模型进行训练,并通过一定的决策方式将8个模型集成为一个协同作业系统,用以预报太阳质子事件24 h内是否发生。本文选取了完整的第23太阳活动周的数据进行实验,并在完整覆盖太阳活动高年和低年的数据上进行了模型构建与测试,取得了80.95%的报准率的同时,将虚报率降低至19.05%,相比现有的预报系统具有较为明显的优势。

1　数据选取

1.1　数据内容

使用包括太阳耀斑、CME、太阳黑子三类数据用于太阳质子事件的短期预报,选取依据和使用的参数如下。

(1)太阳耀斑。太阳耀斑是发生在太阳大气层局部区域的一种剧烈的爆发现象。根据Kahler[8]提出的大耀斑假说,耀斑的大小与高能粒子事件的发生有着密切的联系。Le等[9]的研究也表明,太阳耀斑对于能量大于10 MeV的质子加速有明显贡献。本文所使用的耀斑数据主要包括X射线耀斑等级和耀斑位置两项参数。

(2)CME。CME 是日冕磁场以及积聚在日冕中的等离子体团在较短时间内向行星际爆发的现象。根据 Kahler 等[10] 的研究, CME 与太阳质子事件中高能粒子的峰值通量有很强的相关性。Wang 等[11] 也曾提出, 速度大且源区位置距离日面上连接地球磁力线足点近的全晕 CME 极易引发太阳质子事件。本文所使用的 CME 数据主要包括中心位置角、角宽度、线速度和加速度四项参数。

(3)太阳黑子。太阳黑子是太阳光球层表面深色的区域, 是日面强磁场的产物, 与耀斑和 CME 都具有密切的联系。Mariachiara 等[12] 和 Aoife 等[13] 曾采用太阳黑子相关数据进行耀斑预测。Howard[14] 曾提出太阳黑子处的强磁场与 CME 的形成紧密相关。Bai 等[15] 的研究也表明了太阳黑子与太阳质子事件之间较强的相关性。本文使用的太阳黑子数据主要包括黑子群的磁极性分类、McIntosh 分类[16]、黑子群位置、黑子群面积、黑子数目以及经度范围 6 项参数。

1.2 数据范围与规模

本文所使用的太阳质子事件、太阳耀斑和 CME 的数据来源于美国协作数据分析工作室(CDAW), 太阳黑子的数据来源于美国国家地球物理数据中心(NGDC)官方网站。为了使预报系统全面地学习太阳活动高年和太阳活动低年中太阳质子事件的活动规律, 选用了一个完整的太阳活动周——第 23 个太阳活动周(1997－2007 年)的数据。该选择参考了 Xu 等[17] 对太阳活动周的划分。该活动周内有记录的太阳质子事件共 105 次, 由于设备的故障造成数据残缺, 有 25 次事件未被使用。数据的总量为 654 条, 每条数据由相关联的 CME、太阳耀斑和太阳黑子的参数组成。数据的关联工作由 CDAW 完成。其中有 80 条数据对应了太阳质子事件的发生, 574 条无太阳质子事件发生。

2 实验方法

选取在多个领域中广泛使用的机器学习模型, 采用太阳质子事件预报中常用的评价指标, 以交叉验证的方式计算每个模型的指标并依此为模型赋予权重, 然后通过加权投票的方式将其集成为一个模型。集成模型能够综合所有单一模型的优点, 拥有更强的预报能力。

2.1 模型选用

本质上, 太阳质子事件的短期预报(发生与否)是机器学习领域中常见的二分类问题。近些年来, 一些机器学习分类模型(下称模型)在太阳质子事件预报中得到了一定的应用, 但是仅限于单个模型的使用, 普遍存在虚报率较高的问题。表 1 给出了本文选用的用于太阳质子事件的短期预报 8 种模型。每个模型经过训练后, 对于一条输入数据, 会预报 24 h 内发生质子事件的概率, 当概率大于 50% 时, 认为 24 h 内会有太阳质子事件发生。

2.2 模型评价

2.2.1 评价指标

模型对太阳质子事件的预报能力需要一定的指标来进行衡量。目前太阳质子事件短期预报中广泛采用的评价指标包括报准率($R_{\rm pd}$)和虚报率($R_{\rm fa}$), 即

$$R_{\rm pd}=\frac{N_{\rm TP}}{N_{\rm TP}+N_{\rm FN}}, \tag{1}$$

$$R_{\rm fa}=\frac{N_{\rm FP}}{N_{\rm TP}+N_{\rm FP}}. \tag{2}$$

其中: $N_{\rm TP}$ 为真正例(True Positive,TP), 表示将正类(发生了太阳质子事件的数据)正确预测为正类的数量; $N_{\rm FP}$ 为假正例(False Positive, FP), 表示将负类(未发生太阳质子事件的数据)错误预测为正类的数量; $N_{\rm FN}$ (False Negative, FN) 为假负例, 表示将正类错误预测为负类的数量。

但是单独使用报准率和虚报率两项指标之一, 并不能完全反映一个模型对质子事件的预测能力。因此引入了 F_1 值作为综合评价指标, 其计算方式如下:

表 1 本文使用的 8 种机器学习模型
Table 1 Machine learning models applied in this study

模型名称	简称	引用文献
Logistic Regression	LR	Ref. [18]
Decision Tree	DT	Ref. [19]
Random Forest	RF	Ref. [20]
Adaptive Boosting	AdaBoost	Ref. [21]
Gradient Boosting	GBDT	Ref. [22]
Extreme Gradient Boosting	XGBoost	Ref. [23]
K-Nearest Neighbour	KNN	Ref. [24]
Gaussian Naïve Bayes	GNB	Ref. [25]

$$F_1 = 2 \times \frac{[R_{\rm pd}(1-R_{\rm fa})]}{[R_{\rm pd}+(1-R_{\rm fa})]}. \tag{3}$$

2.2.2　评价方法——交叉验证

交叉验证是用于评价一种机器学习模型是否适用于当前应用场景(在本文中是太阳质子事件短期预报)的实验方法。交叉验证的第一步即是将数据集按照特定规律划分。根据机器学习领域中的惯例,结合本文数据集规模,表 2 给出了本文数据集的划分方式。

(1)将数据集按照约 3:1 的比例划分为训练集和测试集。这种划分并非是随机划分,而是采用时序分层划分的方式。首先将数据按照时间顺序排列,取前 4 条出列,从中随机取 3 条划入训练集,剩余 1 条划入测试集。重复此过程,直至整个数据集被划分完毕。采用这种划分方式,可以使太阳活动高年和太阳活动低年的数据被均匀地分配到训练集和测试集中,使模型对于太阳高年和太阳活动低年的太阳质子事件均有较强的预报能力。

(2)将训练集按照 4:1 的比例进一步划分为交叉训练集和交叉验证集。需要注意的是:在训练集的划分中,交叉训练集和交叉验证集不是固定的,而是将训练集平均分为 5 个子集后随机取 4 个子集合并作为交叉训练集,剩余 1 个子集作为交叉验证集,共有 5 种不同的组合。这里的 5 等分同样采用时序分层的划分方式,将数据按照时间顺序排列后,取前 5 条出列,并将其随机划分到 5 个子集中。重复此过程,直至整个数据集被划分完毕。因此每种组合中都等比例地包含太阳活动高年和太阳活动低年的数据。

将交叉训练集输入到某个模型对其进行训练,并将其对应的交叉验证集输入到训练后的模型中进行预报,则可以获得一组报准率,虚报率和 F_1 值用于评价该模型的预报能力。在训练集 5 种不同的组合上均进行一次实验,称为一次交叉验证。每次交叉验证,都会获得 5 个报准率,虚报率和 F_1 值,以其均值表征模型预报的准确性。

表 2　实验数据的划分方式

Table 2　Dataset division strategy

子数据集		数量	作用	
训练集	交叉训练集	400条 训练集的4/5	交叉验证中训练模型	训练集成模型
	交叉验证集	100条 训练集的1/5	交叉验证中测试模型能力	
测试集		164条	测试集成模型的能力	

2.3　模型集成

模型评价完成后,需要根据评价结果对模型进行集成,以综合利用多个模型的优点。由于 F_1 值综合了一个模型报准率与虚报率,更能代表一个模型的预报能力,因此选用 F_1 值作为每个模型的权重用来对模型进行集成。集成采用了一种加权投票的方式。

设定一条数据输入集成模型后预报所得的输出(即质子事件发生的概率)为$\hat{p}$,则有

$$\hat{p} = \sum_{i=1}^{N}(w_i p_i). \tag{4}$$

其中,N为单个模型的数量,w_i为单个模型的权重,p_i为单个模型的输出。预报阈值设置为 0.5,若$\hat{p} > 0.5$则认为短期内会发生太阳质子事件;反之则认为不会。

为了使集成模型输出的 $\hat{p}$ 取值范围在 0 和 1 之间,需要对每个模型的 F_1 值做归一化处理得到权重

$$w_i = F_{1i} / \sum_{j=1}^{N} F_{1j}. \tag{5}$$

集成后的模型首先使用训练集数据进行训练,训练得到的模型即是最终的预报模型。使用测试集对其预报能力进行测试。测试集则是从未参与过预报模型训练的数据,预报模型对其未知,性质等同于生产环境中新到来的数据,因此可以很好地用来测试集成模型的能力。

3　实验结果

3.1　交叉验证结果

根据表 2 的数据集划分,在训练集上对选用的 8 种模型进行交叉验证后得到的报准率,虚报率和 F_1 值的均值列于表 3。

3.2　模型集成结果

获取到 F_1 值的均值之后,利用 2.3 节中介绍的模型集成方法,将 8 种模型集成为一个最终的预报模型,在训练集上训练后,在测试集上对集成模型测试。

测试集数据共 164 条,有 21 条数据对应了太阳质子事件的发生。表 4 为集成模型对测试集中太阳质子事件的预报结果。其中 17 次被集成模型正确预测,而 2001 年 11 月 17 日,2002 年 8 月 14 日,2003

表 3 八种模型各自在交叉验证中的表现

Table 3 Performance of 8 individual models on cross-validation dataset

模型	报准率	虚报率	F_1值
LR	0.5325	0.3602	0.5753
DT	0.619	0.3810	0.6190
RF	0.5714	0.1429	0.6857
Ada	0.5238	0.2143	0.6286
GBDT	0.6667	0.1765	0.7368
XGB	0.7619	0.3333	0.7111
GNB	1.0000	0.6719	0.4941
KNN	0.5714	0.3333	0.6154

表 4 集成模型对测试集中的太阳质子事件的预报结果

Table 4 SPEs prediction result of the ensemble model on the test set

质子事件发生时间	峰值通量	是否漏报	质子事件发生时间	峰值通量	是否漏报
1998-04-20	1700	否	2002-08-14	26	是
1998-11-07	11	否	2002-08-24	317	否
1999-06-04	64	否	2003-10-26	466	否
2000-11-24	94	否	2003-11-02	30	否
2001-04-02	1110	否	2003-12-02	89	是
2001-04-12	51	否	2004-07-25	2090	否
2001-11-17	34	是	2004-11-01	63	是
2001-11-22	25	否	2005-01-20	1860	否
2002-03-15	13	否	2005-08-22	337	否
2002-04-21	2520	否	2005-09-09	1880	否
2002-07-15	234	否			

年 12 月 2 日, 2004 年 11 月 1 日发生的 4 次太阳质子事件被漏报。由表 4 可以看出, 漏报的 4 次太阳质子事件峰值通量均在 10～100 pfu 之间。

此外, 有 4 条并未对应太阳质子事件发生的数据, 被虚报为有太阳质子事件发生。将测试结果汇总列于表 5, 计算得到报准率为 80.95%, 误报率为 19.05%。

3.3 结果比较

为了直观地说明本文预报方法的性能, 表 6 给出了本文实验结果与美国空军实验室的 PPS 系统, Laurenza 模型, Núñez 等开发的 UMASEP 模型, 以及 Qiuzhen 等方法预报性能的对比情况。由表 6 可以看出, 本文方法在保证 80% 以上报准率的基础上,

表 5 集成模型在测试集上的预报结果

Table 5 Result of the ensemble model on the test set

预报/实际	无SPE	有SPE	总计
无SPE	139	4	143
有SPE	4	17	21
总计	143	21	164

表 6 本文模型与现行其他预报方法的对比

Table 6 Comparison with other current prediction methods

模型	报准率	虚报率	F_1值
PPS	85.71%	50.00%	0.6316
Laurenza模型	63%	42%	0.6040
UMASEP	80.72%	33.99%	0.7263
Qiuzhen模型	80.00%	25.85%	0.7696
本文方法	80.95%	19.05%	0.8085

进一步降低了虚报率, 模型综合表现(F_1 值)有了明显的提升。

4 结语

根据上述的实验与结果进行分析, 本文在以下几个方面取得了一定成果。

(1)提出了一种基于集成学习的太阳质子事件短期预报的方法。该方法选用了 8 种广泛使用的机器学习模型, 通过交叉验证评价和加权投票机制获得了一个集成的预报模型。

(2)本文模型在对太阳质子事件预报的报准率高达 80.95% 的同时, 进一步将虚报率降低至 19.05%。相比现有的预报方法性能有较大的改善。

(3)本文使用了一个完整的太阳活动周的数据, 并且通过时序分层划分的方式, 保证了模型训练和模型测试的数据中都包括了太阳活动高年与太阳活动低年的数据。因此, 本文模型对于太阳活动高年和低年的太阳质子事件均具有较好的预报能力。

本文的集成预报模型在太阳质子事件短期预报任务中拥有较强的能力, 如果能够将该集成模型进行工程化, 形成功能完备的业务预报系统, 并应用于实际航天任务中, 则能更加有效地进行航天任务的规划, 规避或减少对航天器与航天员的伤害。此外, 本文中仅应用了太阳黑子、太阳耀斑、CME 相关数据

进行太阳质子事件短期预报。然而太阳质子事件的形成是一个极为复杂的物理过程。高能粒子的产生，粒子的加速机制与行星际运输过程等，均受到日地空间多重因素影响。近地 X 射线背景、总太阳辐照度、太阳射电通量以及射电噪暴等现象都与太阳质子事件的形成密切相关。相关输入数据集的选择、扩充和完善，则有机会获得更为精准的太阳质子事件短期预报模型。这些都将是下一步研究的方向。

致谢　国家科技资源共享服务平台国家空间科学数据中心(http://www.nssdc.ac.cn)提供计算服务,美国协作数据分析工作室(CDAW)提供太阳耀斑与 CME 相关数据，美国国家地球物理数据中心(NGDC)提供太阳黑子相关数据以及太阳质子事件列表。

参考文献

[1] ARAN A, SANAHUJA B, LARIO D. SOLPENCO: a solar particle engineering code[J]. *Advances in Space Research*, 2006, **37**(6): 1240-1246

[2] LUHMANN J G, LEDVINA S A, ODSTRCIL D, *et al*. Cone model-based SEP event calculations for applications to multipoint observations[J]. *Advances in Space Research*, 2010, **46**(1): 1-21

[3] KAHLER S W, CLIVER E W, LING A G. Validating the Proton Prediction System (PPS)[J]. *Journal of Atmospheric and Solar-Terrestrial Physics*, 2007, **69**(1/2): 43-49

[4] LAURENZA M, CLIVER E W, HEWITT J, *et al*. A technique for short-term warning of solar energetic particle events based on flare location, flare size, and evidence of particle escape[J]. *Space Weather*, 2009, **7**(4): S04008

[5] NÚÑEZ M. Predicting solar energetic proton events ($E >$ 10 MeV)[J]. *Space Weather*, 2011, **9**(7): S07003

[6] HUANG X, WANG H N, LI L P. Ensemble prediction model of solar proton events associated with solar flares and coronal mass ejections[J]. *Research in Astronomy and Astrophysics*, 2012, **12**(3): 313-321

[7] ZHONG Q Z, WANG J J, MENG X J, *et al*. Prediction model for solar energetic proton events: analysis and verification[J]. *Space Weather*, 2019, **17**(5): 709-726

[8] KAHLER S W. The role of the big flare syndrome in correlations of solar energetic proton fluxes and associated microwave burst parameters[J]. *Journal of Geophysical Research: Space Physics*, 1982, **87**(A5): 3439-3448

[9] LE Guiming, WANG Hongyan, BAI Tienan. Relationship between solar proton events and the associated solar flares[J]. *Chinese Journal of Space Science*, 2018, **38**(4): 437-443 (乐贵明, 王鸿雁, 白铁男. 太阳质子事件与太阳耀斑的关系[J]. 空间科学学报, 2018, **38**(4): 437-443)

[10] KAHLER S W, SHEELEY N R JR, HOWARD R A, *et al*. Associations between coronal mass ejections and solar energetic proton events[J]. *Journal of Geophysical Research: Space Physics*, 1984, **89**(A11): 9683-9693

[11] WANG Cong, CUI Yanmei, AO Xianzhi, *et al*. Relationship of halo CME and solar proton events[J]. *Chinese Journal of Space Science*, 2018, **38**(1): 9-18 (王聪, 崔延美, 敖先志, 等. 全晕CME与太阳质子事件的关系[J]. 空间科学学报, 2018, **38**(1): 9-18)

[12] FALCO M, COSTA P, ROMANO P. Solar flare forecasting using morphological properties of sunspot groups[J]. *Journal of Space Weather and Space Climate*, 2019, **9**: A22

[13] MCCLOSKEY A E, GALLAGHER P T, BLOOMFIELD D S. Flare forecasting using the evolution of McIntosh sunspot classifications[J]. *Journal of Space Weather and Space Climate*, 2018, **8**: A34

[14] HOWARD T. Coronal Mass Ejections[M]. New York: Springer, 2011

[15] BAI Tienan, LE Guiming, ZHAO Haofeng. Statistical properties of solar proton events during solar cycle 23 and 24[J]. *Chinese Journal of Space Science*, 2017, **37**(6): 649-658 (白铁男, 乐贵明, 赵浩峰. 第23至24周太阳质子事件的统计特征[J]. 空间科学学报, 2017, **37**(6): 649-658)

[16] MCINTOSH P S. The classification of sunspot groups[J]. *Solar Physics*, 1990, **125**(2): 251-267

[17] XU Zhenzhong, WANG Weimin, ZHANG Ren, *et al*. Characteristic analysis of ionosphere TEC at Wuhan station during 23 rd solar cycle[J]. *Chinese Journal of Space Science*, 2013, **33**(1): 28-33 (徐振中, 王伟民, 张韧, 等. 第23太阳活动周武汉站电离层TEC特征分析[J]. 空间科学学报, 2013, **33**(1): 28-33)

[18] HARRISON E, RIINU P. Logistic regression[M]// R for Health Data Science. New York: Chapman and Hall/CRC, 2020.

[19] WANG P H, TU Y S, TSENG Y J. PgpRules: a decision tree based prediction server for P-glycoprotein substrates and inhibitors[J]. *Bioinformatics*, 2019, **35**(20): 4193-4195

[20] FEZAI R, BOUZRARA K, MANSOURI M, *et al*. Random forest-based nonlinear improved feature extraction and selection for fault classification[C]//Proceedings of 2021 18th International Multi-Conference on Systems, Signals & Devices (SSD). Monastir: IEEE, 2021: 601-606

[21] FREUND Y, SCHAPIRE R E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. *Journal of Computer and System Sciences*, 1997, **55**(1): 119-139

[22] BIAU G, CADRE B, ROUVIÈRE L. Accelerated gradient boosting[J]. *Machine Learning*, 2019, **108**(6): 971-992

[23] CHEN T Q, GUESTRIN C. XGBoost: a scalable tree boosting system[C]//Proceedings of the 22 nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: ACM, 2016: 785-794

[24] ZHU H J, ZHU L H. Encrypted network behaviors identification based on dynamic time warping and k-nearest neighbor[J]. *Cluster Computing*, 2019, **22**(2): 2571-2580

[25] PÉREZ A, LARRAÑAGA P, INZA I. Supervised classification with conditional Gaussian networks: increasing the structure complexity from naive Bayes[J]. *International Journal of Approximate Reasoning*, 2006, **43**(1): 1-25

(责任编辑：孙伟英)