控制与决策
Control and Decision

蓄意攻击样本有限不均衡下运输系统关键危险源识别

杨黎霞， 许茂增， 陈仁祥， 吴昊年

引用本文:
杨黎霞, 许茂增, 陈仁祥, 等. 蓄意攻击样本有限不均衡下运输系统关键危险源识别[J]. 控制与决策, 2022, 37(2): 464−472.

在线阅读 View online: https://doi.org/10.13195/j.kzyjc.2020.1143

蓄意攻击样本有限不均衡下运输系统关键危险源识别

杨黎霞[1,2], 许茂增[1], 陈仁祥[3†], 吴昊年[3]

(1. 重庆交通大学 经济与管理学院，重庆 400074；2. 重庆广播电视大学 管理学院，重庆 400052；
3. 重庆交通大学 机电与车辆工程学院，重庆 400074)

摘 要: 针对蓄意攻击样本有限不均衡而引起无法有效识别关键危险源少数类样本的问题,提出多分类器集成加权均衡分布适配的关键危险源识别方法. 首先,在保证少数类样本被充分选择的前提下随机抽取多数类样本,构成源域多样本训练集合,在目标域上直接预测伪标签并给样本赋予不同的权重,让少数类样本可以得到充分的训练;然后,训练源域样本集的分类器,经过多次迭代优化目标域伪标签并更新权重矩阵;最后,通过多分类器集成的策略将筛选出的基分类器集成为强分类器,采用宏平均和微平均两个评价指标来评价分类器的识别性能. 利用全球恐怖主义数据库(GTD)中的数据进行实验验证,实验结果表明所提出方法在保证了整体精度的同时能有效识别少数类样本.

关键词: 运输系统；蓄意攻击；关键危险源；样本有限不均衡；多分类器集成；智能识别

中图分类号: X951 **文献标志码:** A

DOI: 10.13195/j.kzyjc.2020.1143 **开放科学(资源服务)标识码(OSID):**

引用格式: 杨黎霞,许茂增,陈仁祥,等. 蓄意攻击样本有限不均衡下运输系统关键危险源识别[J]. 控制与决策, 2022, 37(2): 464-472.

Intelligent identification of critical hazard sources in transport system with deliberate attack sample finite unbalance

YANG Li-xia[1,2], *XU Mao-zeng*[1], *CHEN Ren-xiang*[3†], *WU Hao-nian*[3]

(1. School of Economics and Management，Chongqing Jiaotong University，Chongqing 400074，China；2. School of Management，Chongqing Radio and Television University，Chongqing 400052，China；3. School of Mechatronics and Vehicle Engineering，Chongqing Jiaotong University，Chongqing 400074，China)

Abstract: In order to solve the problem that samples of minority class of critical risk sources can't be effectively identified due to the deliberate attack samples finite unbalance, a multi-classifier ensemble weighted balanced distribution adaptive method for critical risk sources identification is proposed. Firstly, ensuring that the minority samples are fully selected, the source domain multi sample training set is obtained by random sampling, and different initial weights are given to the samples to fully train the minority samples. Then, the classifier of the sample set in the source domain is trained, and the pseudo label of the target domain is optimized and the weight matrix is updated after many iterations. Finally, the selected base classifiers are integrated into strong classifiers through the strategy of multi classifier integration, and the recognition performance of classifiers is evaluated by macro average and micro average evaluation indexes. The global terrorism database (GTD) data is used to verify the proposed method, which can effectively identify a small number of samples while ensuring the overall accuracy.

Keywords: transportation system；deliberate attack；critical hazard sources；sample finite unbalance；multi-classifier integration；intelligent recognition

0 引 言

近年来,蓄意攻击行为频繁发生,给国际社会的安全问题带来了巨大的挑战. 9·11事件后,社会服务系统的安全问题引起高度重视[1]. 交通运输系统作为一个开放的公共场所,积聚了大量的人流和车流在相对封闭的空间里移动[2],因此常成为蓄意攻击的目标. 运输系统一旦被蓄意攻击将引起重大的人员伤亡和财产损失,同时会导致运输系统的级联失效[3],由此可见运输系统的安全问题尤为重要,需要极大的关注[4].

收稿日期: 2020-08-17；**录用日期:** 2020-12-03.
基金项目: 国家自然科学基金项目(71471024).
责任编委: 李登峰.
†通讯作者. E-mail: manlou.yue@126.com.

蓄意攻击是按照一定的策略进行针对性地攻击[5]. 由全球恐怖主义数据库(global terrorism database, GTD)[6]统计数据可知,蓄意攻击的攻击方式包含了暗杀、劫持、绑架、路障事件、轰炸/爆炸、武装突袭、徒手攻击、设施/基础设施攻击8大类,8类攻击方式对应着不同的关键危险源. 运输系统可能遭遇一种或多种关键危险源的攻击,当运输系统被袭击后,针对不同的危险源所采用的应急救援有巨大的差异. 如当运输系统遭遇轰炸/爆炸后,搜爆和排爆工作可以有效防止第2次爆炸,同时需对爆炸现场进行分区管理和制定疏散计划直至医学救援团队到来[7-8]. 而徒手攻击中的生化攻击与轰炸/爆炸袭击的特性有较大差别,生化危险源具有隐蔽性、扩散性和传染性,防范生化危险源的重点是前期的监测,生化攻击发生后的应急救援需要全面公共卫生反应[9]. 由此可见,如何针对不同的关键危险源进行快速、高效和准确地识别,预判出关键危险源为蓄意攻击事件预警和应对提供数据支撑,是交通反恐的关键问题.

针对蓄意攻击的关键危险源识别问题,相关学者已展开研究. 如: Nizamani等[10]基于GTD的文本数据,就决策树、朴素贝叶斯和支持向量机3种分类方法进行了对比分析,研究表明支持向量机能达到合理的准确率,但是运行时间太长,朴素贝叶斯虽速度快而准确率低,相比较决策树的综合表现更好,分类识别率能达到83 %. Sivaraman等[11]针对关键危险源识别提出基于多分类器的集成决策树算法,利用GTD数据进行实例分析,结果表明该算法比单一决策树算法的准确率有显著提高. 肖圣龙等[12]提出基于Spark平台的分布式神经分类算法,采用GTD所提供的数据集验证其算法,结果表明该算法较集成决策树算法提高了识别平均准确率,但部分危险源的识别准确率有所下降. Meng等[13]提出了一种用于预测关键危险源的优化混合分类器算法,结果表明混合分类器的预测精度优于单一分类器. 以上研究工作在各类危险源样本数量相同的条件下取得了较好效果,但尚未涉及各类危险源样本数量有限且不均衡的问题. 蓄意攻击对社会具有破坏性,不能人为进行实验去扩充样本量,致使样本量有限. 蓄意攻击的特点是有策略的攻击,攻击者会根据不同危险源的使用条件和难度选择性地采用成功率高的类别,因此每类危险源被采用的概率差异较大,导致8类危险源的样本数量呈现出严重不均衡,以GTD中1970～2017年数据为例,其中多数样本(bombing explosion)有14 650条数据,而少数类样本(barricade incident)仅有74条数据. 由于多数类样本会模糊少数类样本的边界,在某些存在类别重叠的区域,很难有效地将少数类样本与多数类样本区分开来,容易造成分类器对少数类样本识别度下降[14]. 这种样本数量有限且不均衡增加了危险源识别的难度,只有将淹没在多数类样本中的少数类样本有效识别,才能保证每一类样本识别的准确性.

要解决样本有限不均衡下蓄意攻击关键危险源识别的问题,必须解决两大难题: 1) 多数类样本淹没了少数类的样本所含有的信息,使得模型得不到有效训练. 2) 弱分类器识别的结果更倾向于多数类致使少数类的识别错误率高[15]. 迁移学习利用少量有标记数据,能从不同关键危险源数据中挖掘出有价值的信息共同训练模型,它打破了传统机器学习独立同分布的假设,可以实现跨任务、跨领域的学习. 如: 么素素等[16]利用基于级联模型的集成迁移学习算法(TrAdaBoost)以及陈琼等[17]采用不均衡数据迁移学习算法解决了二分类的样本不均衡问题. Wang等[18]提出加权均衡分布适配方法(weighted balanced distribution adaptation, W-BDA),该方法是用类先验精确地逼近目标类别条件分布,从而提升训练分类器的性能,可以解决样本数量不均衡的问题,但是当样本数据极度不均衡时表现不佳. 为解决样本数据极度不均衡下的识别问题,陈仁祥等[15]在W-BDA基础上进行多分类集成,形成了多分类器集成加权均衡分布适配的滚动轴承寿命阶段识别方法,效果良好.

综上所述,针对蓄意攻击数据的样本有限不均衡引起的无法有效识别关键危险源少数类样本的问题,提出了多分类器集成加权均衡分布适配的关键危险源识别方法. 首先,保证少数类样本得到有效训练,在充分选择少数类样本的前提下随机抽取多数类样本形成源域多样本训练集,充分训练后提高少数类的权重,同时直接预测没有标签的目标域测试样本得到多个目标域伪标签;然后,利用类先验概率逼近条件分布概率的策略构建基分类器,采用源域多样本训练集和目标域标记样本集训练得到对应样本的识别结果;最后,根据多分类器集成思路,有效集成基分类器信息构成强分类器完成多分类识别,可以解决样本有限不均衡下蓄意攻击关键危险源识别问题. 采用GTD收录数据进行验证,结果表明本文所提出方法可行且有效.

1　多分类器集成加权均衡分布适配算法

1.1　多分类器集成

在解决样本有限不均衡时,加权均衡分布适配需充分训练少数类样本才能构建较好的训练模型,与此

同时,弱分类器在识别时结果更倾向于多数类,而对于少数类不能较好地进行识别. 为了更好地解决样本有限不均衡下关键危险源的识别问题,构建如图1所示的源域多样本训练集,并训练相应的基分类器,再采用弱分类器集成的方式来获得最终结果,从而提升分类器的整体性能. 加权最近邻(weight K nearest neighbors, WKNN)分类器不需要如支持向量机、决策树等分类算法的复杂训练过程,仅用数据局部信息即可得到分类结果,计算效率高、易于实现,且相对于最近邻(K nearest neighbors, KNN)分类器受邻域大小影响小、鲁棒性更好,因此以WKNN作为基分类器.

WKNN的主要算法如下.

step 1: 利用欧氏距离计算测试样本x_t与每个训练样本x_i之间的距离

$$d(x_t,x_i)=\sqrt{\sum_{j=1}^{m}\|x_{tj}-x_{ij}\|^2}. \tag{1}$$

根据$d(x_t,x_i)$大小,从训练集X中找出x_t的$k+1$个近邻样本$x_{t,1},x_{t,2},\dots,x_{t,k+1}$.

step 2: 从k+1个近邻样本中选择与x_t距离最大的样本设为$x_{t,k+1}$,相应的距离为$d(x_t,x_{t,k+1})$,利用$d(x_t,x_{t,k+1})$对其他k个近邻样本与x_t的距离进行标准化,即

$$D(x_t,x_i)=\frac{d(x_t,x_i)}{d(x_t,x_{t,k+1})},\ i=1,2,\dots,k. \tag{2}$$

step 3: 对标准化后的距离$D(x_t,x_i)$,利用高斯核函数将其转化为x_t与x_i的同类概率

$$p(x_i|x_t)=\frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{D(x_t,x_i)}{2}\Big). \tag{3}$$

step 4: 根据x_t与k个近邻样本的同类概率$p(x_i|x_t)$,求出x_t为类别$l_i(i=1,2,\dots,r)$的后验概率

$$P(l_i|x_t)=\frac{\sum\limits_{x_i\in X}\begin{cases}0,\ l_i\neq l_t;\\ p(x_i|x_t),\ l_i=l_t\end{cases}}{\sum\limits_{x_i\in X}p(x_i|x_t)}. \tag{4}$$

则可获得最有可能的分类结果

$$\mathrm{KNN}(x_t)=\arg\max_{l_i\in L}\{P(l_i|x_t)\}. \tag{5}$$

WKNN以各近邻样本与测试样本的相似程度对近邻样本赋予不同的权重,使得测试样本的分类结果更加接近于相似程度更高的训练样本,从而会提高识别精度和削弱对k值选择的敏感性,使得识别结果的精度和鲁棒性更好.

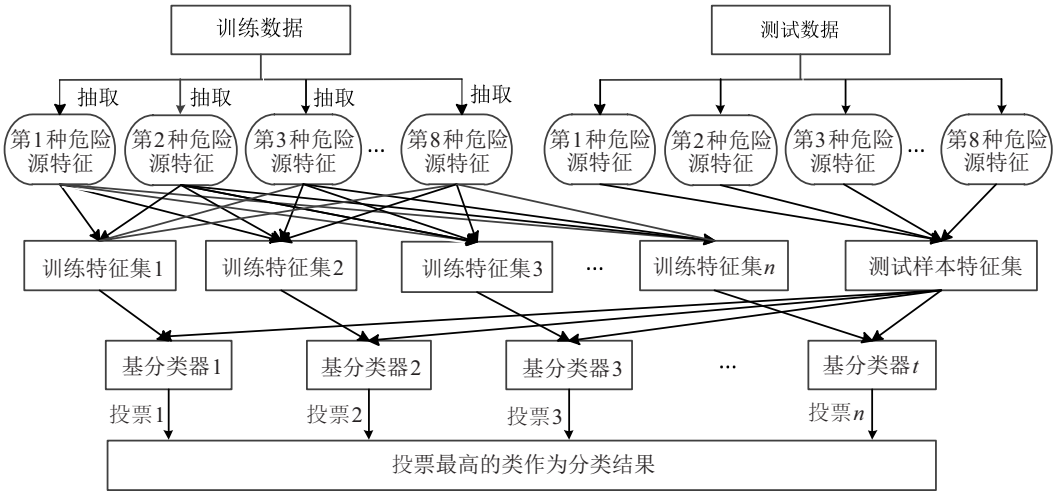


图 1 多分类器集成

假设X_s为M个不同危险源的特征样本,危险源为$\varphi_i,\forall_i\in\Lambda=\{1,2,\dots,M\}$,则$\varphi_1$代表第1种危险源特征样本. 针对$X_s$不同危险源的样本中的多数类样本采用无重复随机抽样,抽取样本数为$n,n\leqslant\min(\varphi_1,\varphi_2,\dots,\varphi_M)$,构成一个源域样本训练集$x_s$,即少数类样本为抽取多数类样本的最小基数. 当少数类样本数量确定后,以少数类样本最小基数抽取多数类,确保每次均会被抽取到少数类样本和对其进行充分训练. 然后,对多数类样本进行k次抽样直至抽完全部样本,构成k个源域单样本训练集$\{x_{s1},x_{s2},\dots,x_{sk}\}$. 随机抽取多数类样本得到源域多样本训练集的前提是要充分选择少数类样本,充分训练少数类样本,增加少数类样本的权重,从而得到多个目标域伪标签. 再通过下式:

$$\varepsilon=\sum_{k=1}^{n}\frac{w_{ck}|h_c(x_{sk}-y_{si})|}{\sum\limits_{k=1}^{n}w_{ck}} \tag{6}$$

计算得出伪标签分类器的错误率,将错误率最小的作

为基分类器$h_c(t=1,2,\ldots,k)$,假设第j类的标记为$w_j(j=1,2,\ldots,M)$,第c个分类器对第j类样本的预测值为$h_c(j)$.

采用多分类器集成的策略,将多个弱分类器识别结果代入下式进行一致性投票,得到最终的识别结果:

$$f_E(x_{sk})=\arg\max(t|h_c(j)=\varphi_i). \tag{7}$$

1.2　加权均衡分布适配

参数设置为: x_i为源域特征样本集,y_i为样本标签向量,源域为$D_S=\{x_{si},y_{si}\}_{i=1}^n$,无标记目标域为$D_t=\{x_{tj}\}_{j=1}^m$. 假设源域和目标域具有相同的特征空间$x_s=x_t$,相同的类别空间$y_s=y_t$,不同的边缘分布$p_s(x_s)\neq p_t(x_t)$和不同条件的概率分布$p(y_s|x_s)\neq p(y_t|x_t)$.

加权均衡分布适配的基本原理是,寻找到一个变换矩阵A,经过变换后,使得源域条件分布$p(y_s|A^{\mathrm{T}}x_s)$和目标域条件分布$p(y_t|A^{\mathrm{T}}x_t)$具有最小的最大均值差[15]. 由于目标域里不包含类别向量y_t,无法直接得到目标域的条件分布,采用$p(x_t|y_t)$来近似$p(y_t|x_t)$,因此源域和目标域的条件分布差异[19]为

$$\begin{aligned}&\|p(y_s|x_s)-p(y_t|x_t)\|_H^2=\\&\left\|\frac{p(y_s)}{p(x_s)}p(x_s|y_s)-\frac{p(y_t)}{p(x_t)}p(x_t|y_t)\right\|.\end{aligned} \tag{8}$$

其中隐含假设每个类别在源域和目标域中的概率是相似的.显而易见,在样本有限不均衡的情况下这个假设不合理, 因此加权均衡分布适配假设$p(x_s)$和$p(x_t)$不变,利用类先验概率逼近条件分布概率来避免式(8)中条件分布散度的计算. 将结果通过矩阵技巧和正则化形式化为

$$\begin{aligned}&\min\left(A^{\mathrm{T}}X\left((1-\mu)M_0+\mu\sum_{c=1}^{c}W_c\right)X^{\mathrm{T}}A\right)+\lambda\|A\|_F^2;\\&\text{s.t. }A^{\mathrm{T}}XHX^{\mathrm{T}}A=I,\ 0\leqslant\mu\leqslant1.\end{aligned} \tag{9}$$

其中: 正则化项$\|g\|_F^2$的Frobenius系数为λ; 平衡因子$\mu\in[0,1]$用来适配源域和目标域的分布; A为变换矩阵; x_s和x_t组成矩阵X; $\boldsymbol{I}\in R^{(n+m)\times(n+m)}$为单位矩阵; $\boldsymbol{1}$为1的列向量; $H=\boldsymbol{I}-(1/n)\boldsymbol{1}$为中心矩阵; M_0为MMD矩阵; W_c为权重矩阵.

$$(M_0)_{ij}=\begin{cases}\dfrac{1}{n^2},\ x_i,x_j\in D_s;\\[2mm]\dfrac{1}{m^2},\ x_i,x_j\in D_t;\\[2mm]-\dfrac{1}{mn},\ \text{otherwise};\end{cases} \tag{10}$$

$$(W_c)_{ij}=\begin{cases}\dfrac{p(y_s^{(c)})}{n_c^2},\ x_i,x_j\in D_s^{(c)};\\[2mm]\dfrac{p(y_t^{(c)})}{m_c^2},\ x_i,x_j\in D_t^{(c)};\\[2mm]-\dfrac{\sqrt{p(y_s^{(c)})p(y_t^{(c)})}}{m_cn_c},\ \begin{cases}x_i\in D_s^{(c)},\ x_j\in D_t^{(c)};\\x_i\in D_t^{(c)},\ x_j\in D_s^{(c)};\end{cases}\\0,\ \text{otherwise}.\end{cases} \tag{11}$$

式(11)中$P(y_s^{(c)})$和$P(y_t^{(c)})$分别为源域和目标域中类C对应的类先验,引入拉格朗日乘子$\Phi=(\Phi_1,\Phi_2,\ldots,\Phi_d)$,将式(9)和(10)代入(11)得到如下拉格朗日函数:

$$\begin{aligned}&\left(X\left((1-\mu)M_0+\mu\sum_{c=1}^{c}W_c\right)X^{\mathrm{T}}+\lambda I\right)A=\\&XHX^{\mathrm{T}}A\Phi.\end{aligned} \tag{12}$$

在式(9)的约束条件下,求解出式(12)的目标函数,可以得到用以构建基分类器的最佳映射变换矩阵A.

2　多分类器集成加权均衡分布的关键危险源识别流程与结果评价

2.1　识别流程

参照上文所述,图2所示为本文所提出方法的识别流程,主要包括以下步骤.

step 1: 构建训练样本. 在充分选择少数类样本的前提下,随机抽样多数类样本,构成源域多样本训练集,让少数类样本得到充分的训练.

step 2: 设置权重系数. 设置初始化参数,构建多分类集成加权均衡分布学习网络,构造初始化权重矩阵. 在目标域上预测测试样本的伪标签,进行多次迭代,更新目标域标签和权重矩阵,采用类先验概率逼近条件分布概率策略,选择伪标签分类器的错误率最小的构建基分类器.

step 3: 多分类器集成. 采用多分类器集成的策略,将多个弱分类器识别结果进行一致性投票,获得多分类器集成输出的识别结果.

2.2　结果评价

准确率是度量一般分类器性能的重要指标,但在样本有限不均衡的情况下,仅单一的采用准确率不能有效地衡量分类器的性能. 例如不平衡样本数为100个(其中多数类样本99个,少数类样本1个),当多数类样本99个被识别,识别准确率达到99 %,将会掩盖少数类样本未被识别的事实. 为此本文引入混淆矩阵

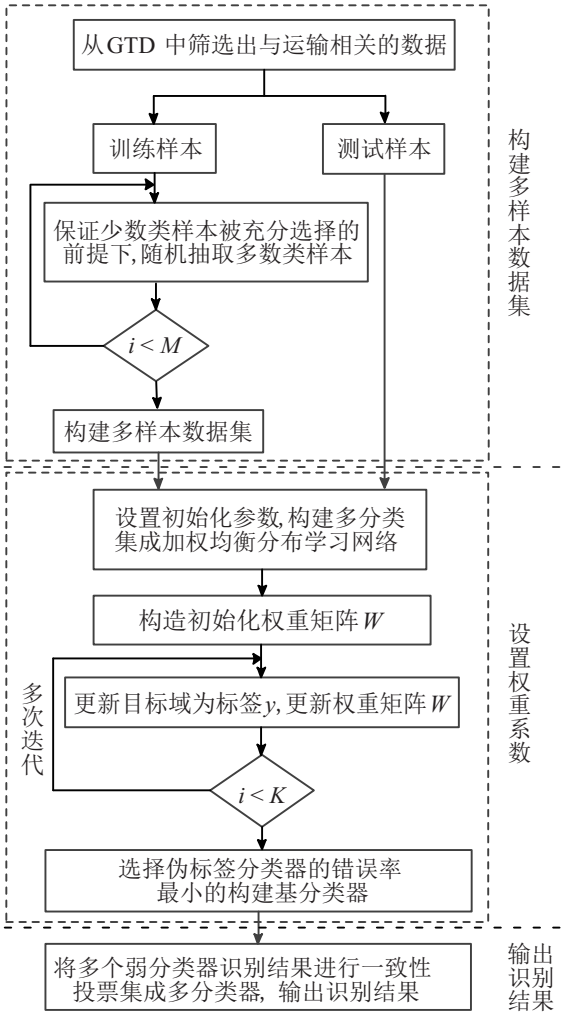


图 2 关键危险源识别流程

用于衡量分类器的性能,如表1所示.基于表1,可以生成常用的衡量二分类结果的评价指标:查准率(Pr)、召回率(Re)以及F分数($F_{\rm score}$)[14].

表 1 混淆矩阵

混淆矩阵		识别类别	
		RIGHT	WRONG
实际类别	RIGHT	RP	WN
	WRONG	WP	RN

$$\Pr=\frac{\rm RP}{\rm RP+WP},\tag{13}$$

$$\mathrm{Re}=\frac{\rm RP}{\rm RP+WN},\tag{14}$$

$$F_{\rm score}=\frac{(1+\beta^2)\times\Pr\times\mathrm{Re}}{(\beta^2\times\Pr)+\mathrm{Re}}.\tag{15}$$

当$F_{\rm score}\geqslant 0.6$时,说明分类结果良好. 同时,将宏平均和微平均用于衡量多类别识别任务中分类器性能. 宏平均(Macro_F)首先统计每一类的指标值,然后计算每一类的算术平均值,重点关注每一类别的识别效果,Macro_F计算公式如下:

$$\mathrm{Macro}_F=\frac{(1+\beta^2)\times\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\frac{\mathrm{RP}_i}{\mathrm{RP}_i+\mathrm{WP}_i}\times\frac{1}{n}\sum_{i=1}^{n}\frac{\mathrm{RP}_i}{\mathrm{RP}_i+\mathrm{WN}_i}}{\left(\beta^2\times\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\frac{\mathrm{RP}_i}{\mathrm{RP}_i+\mathrm{WP}_i}\right)+\dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\frac{\mathrm{RP}_i}{\mathrm{RP}_i+\mathrm{WN}_i}}.\tag{16}$$

微平均(Macro_F)首先统计数据集中不分类别的每一个实例,然后构建全局混淆矩阵计算得到指标,重点关注样本整体的识别效果,本文中β取值为1,即Macro_F等于准确率. Macro_F计算公式如下:

$$\mathrm{Macro}_F=\frac{(1+\beta^2)\times\dfrac{\displaystyle\sum_{i=1}^{n}\mathrm{RP}_i}{\displaystyle\sum_{i=1}^{n}\mathrm{RP}_i+\sum_{i=1}^{n}\mathrm{WP}_i}\times\frac{\displaystyle\sum_{i=1}^{n}\mathrm{RP}_i}{\displaystyle\sum_{i=1}^{n}\mathrm{RP}_i+\sum_{i=1}^{n}\mathrm{WN}_i}}{\left(\beta^2\times\dfrac{\displaystyle\sum_{i=1}^{n}\mathrm{RP}_i}{\displaystyle\sum_{i=1}^{n}\mathrm{RP}_i+\sum_{i=1}^{n}\mathrm{WP}_i}\right)+\dfrac{\displaystyle\sum_{i=1}^{n}\mathrm{RP}_i}{\displaystyle\sum_{i=1}^{n}\mathrm{RP}_i+\sum_{i=1}^{n}\mathrm{WN}_i}}.\tag{17}$$

3 实验与对比分析

3.1 数据预处理

本文数据来源于GTD,由于该数据库收集时间跨度大、蓄意攻击事件描述的复杂性等诸多因素,致使其数据具有不完整、描述重复、不规范、数据类型多样化、数据异常等问题. 在传入模型训练前,需对其进行预处理,包括以下几个步骤.

1)数据筛选.

以1970年~2017年(含1993年)为期,从GTD中筛选与交通相关数据(攻击目标为交通系统、武器为交通工具以及武器为汽车炸弹). 统计数据如表2所示.

表 2 关键危险源分类统计信息

编号	关键危险源	样本数量
1	assassination	602
2	armed assault	1 631
3	bombing explosion	14 650
4	hijacking	409
5	barricade incident	74
6	kidnapping	179
7	infrastructure attack	763
8	unarmed assault	143

2)筛选有效属性.

GTD中蓄意攻击事件每条含有135个属性,其中部分属性解释量小、重复定义、数据缺失严重,需对部分属性剔除[20]. 保留属性包括eventid、iyear、imonth、iday、extended等35个.

3)数据填补.

保留属性仍有部分数据缺失,根据不同属性的特点,采用相应处理方法对缺失值进行填补. 如:利用水经注万能地图将属性latitude和longitude进行填补;用targtype1中各类的子类中被袭频率最高的类来填补targsubtype1中的空白.

4)数据转换.

需要将字符串和日期型的数据转换成数值型. 本文将字符串和日期型的属性通过Excel透视表排序,其序号作为该属性的一个映射值,让其作为输入的源数据[12].

5)数据规范化.

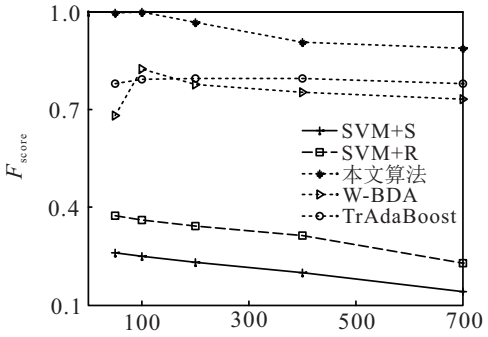
不同属性有不同的量纲,数值间差别较大. 在此采用离差标准化法进行数据规范化处理,将数值映射到[0, 1]间,便于深度神经网络的输入. 转换公式为

$$x^* = \frac{x - x_{\min}}{x_{\max} - x_{\min}}. \tag{18}$$

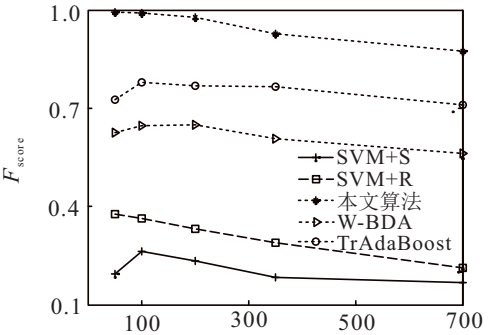
其中: $x_{\max}$为样本数据最大值, $x_{\min}$为样本数据最小值.

3.2　样本数量不平衡比例对识别结果影响分析

为验证本文方法在采样方式上解决数据不均衡情况下识别的优势,选择原始数据8类关键危险源中样本数最少的类别5(barricade incident,样本数74个)作为少数类样本,样本数最多的类别3(bombing explosion,样本数14 650)以及样本数次多的类别2(armed assault,样本数1 631)分别作为多数类,从二分类问题入手,设置多数类(类别3)与少数类(类别5)样本不平衡比例为50:1, 100:1, 200:1, 400:1, 700:1. 设置次多数类(类别2)样本与少数类(类别5)样本不平衡比例为1:10, 1:20, 1:40, 1:70, 1:140. 基于抽样数据,对比分析基于SVM和SMOTE过抽样相结合的方法(SVM+S)、SVM和随机采样(RS)平衡样本相结合方法(SVM+R)、迁移学习TrAdaBoost方法和W-BDA算法的性能. 将整体不平衡样本随机分为10份,其中训练样本8份,测试样本2份. 多数类样本与少数类样本最大比例约为1:200,次多数类样本与少数类样本最大比例约为1:20,因此在这个范围内取不平衡比例,测试样本最大不平衡比例为1:700,最小为1:10. 实验中设置SVM采用宽度为1的高斯核函数,SMOTE算法中$K = 5$, W-BDA迭代次数为$N = 10$,正则化参数为$\lambda = 0.01$. 实验结果如图3所示.



(a)　类别3与类别5各算法性能对比



(b)　类别2与类别5各算法性能对比

图 3　不平衡样本下算法性能对比结果

F_{score}性能随着不均衡数据比例的增大而减少;实验中发现单纯使用随机欠采样方法不能有效提高原有算法性能. 欠采样方式丢弃了很多反例致使分类训练集小于初始训练集,可能造成重要信息的丢失,导致SVM分类器识别结果倾向于多数类. 在图3中SVM+R的方法随着不平衡比例增加,曲线趋势逐次递减; SMOTE过采样算法使用k近邻法,在少数类样本之间进行插值来产生额外的样本. 这样做容易带来两个问题: 1)若选取的少数类样本周围也都是少数类样本,则新合成样本不会提供太多有用信息,使得SVM中远离边界的点对决策边界影响不大; 2)若选取的少数类样本周围全是多数类样本,则这类样本便成了“噪音”,新合成样本会与周围的多数类样本产生大部分重叠,导致SVM分类困难. 图3(a)中随着不平衡比例增加, SVM+S方法F_{score}得分依次下降,而在图3(b)中1:10到1:20过程中得分增加. 通过分析计算过程,图3(b)中SMOTE选取的少数类周围正是少数类样本(类别2与类别5少数类样本较为聚集),所以导致了偶尔的提升,但随着不平衡比例增大其得分总体趋势是下降的.

TrAdaBoost作为迁移学习方法,在不平衡样本

的识别上已有了很好的应用,但TrAdaBoost算法适用于解决基于对称的二分类问题,正负样本同样的权值更新策略,势必造成识别效果不佳. 在图3(b)中1:10到1:20时F_{score}不降反升是因为将两种不平衡样本权值同等看待,使得少数类样本在矩阵更新时增加,多数类样本不变,在少数类样本到达极限以后,F_{score}势必会逐渐下降(类别2与类别5最大比例1:20). 同时, TrAdaBoost辅助数据中往往存在大量冗余数据. 引入辅助样本提高了整体识别准确率,但少数类样本所占比例较小,即便全部识别错误,整体识别准确率仍然很高. 只有在F_{score}的评判标准下才会发现端倪. 随机抽样的W-BDA算法使用两域类先验概率近似逼近条件概率,算法初始就为不均衡样本多数类与少数类赋予了不一样的权重,并且结合随机抽样的方式构建源域多样本训练基分类器,通过基分类器识别结果一致性判别形成强分类器,使得其识别性能大大提升. 这也是在不均衡比例达到1:700时,仍然可以获得高F_{score}得分(F_{score}得分0.89左右)识别结果的原因.

3.3 多类别样本数量不平衡下实验分析

通过上述实验验证了本文方法在与传统不均衡样本解决方法中的优势和可行性,证明了本文方法对于样本绝对不均衡情况识别效果良好,而在关键危险源识别中,至少要同时识别多个样本数量互不均衡的样本情况,二分类方法显得不再适用.

本文方法可以将多类别同时识别,完成多分类任务. 为了对比多分类效果,将上个实验效果比较好的3类算法在多分类任务中再进行对比分析. 其中结合决策树将第N次迭代后得到的多分类器模型h_N作为TrAdaBoost的最终输出,使其可以满足多分类需求. 本节实验仍采用上节中采集的样本集,以验证样本不均衡下多类别分类时本文提出分类器的性能. 实验样本数量及性质如表3所示.

表 3 实验样本信息

样本类别	关键危险源	样本数量
类别1	assassination	602
类别2	armed assault	1 631
类别3	bombing explosion	14 650
类别4	hijacking	409
类别5	barricade incident	74
类别6	kidnapping	179
类别7	infrastructure attack	763
类别8	unarmed assault	143

实验中对8类样本设置两种比例,其中比例1为训练样本与测试样本比例相同,均为602:1 631:14 560:409:74:179:763:143,训练样本占总样本数8成,测试样本为剩余的2成;比例2中将训练样本数量调整为总数据的5成,测试样本比例中各类数量除类别3外急剧减少,并且最大不平衡比例变为1:1 250(类别6与类别3),使训练样本不平衡且测试样本也不平衡,甚至使两样本不平衡比例也不相同,以验证分类器可靠性能. 实验中测试集比例为40:200:2 500:50:30:20:100:50,其中比例1的实验结果如图4所示.

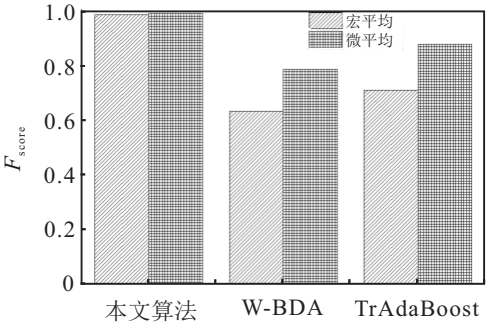


图 4 不平衡样本学习方法性能对比

由图4可知,在训练样本与测试样本不均衡、比例相同的情况下,本文算法具有很好的适应能力,宏平均为0.990 1,微平均为0.994 4. 明显优于未改进前TrAdaBoost算法. 面对比例2的极端条件下,本文方法宏平均得分依旧接近0.8,明显优于对比算法的0.55,微平均(本文代表准确率)也在90 %以上,优于对比算法的80 %. 进一步验证了本文算法可以很好地解决多分类问题,显而易见是本文方法能够较好地解决样本有限不均衡的问题,在保证整体识别精度的前提下,有效识别了少数类样本. 需要说明的是,本文介绍的所有任务都是在工作站(Intel (R) Xeon (R) CPU E5-2 678 v3 @ 2.50 GHz)上完成的,完成本次试验总共用时为65 s,这说明本文算法时间复杂度较低,具有很强的实际应用效果.

3.4 迁移学习对比实验

上个实验中TrAdaBoost也展现了较好的识别能力(F-$\geqslant$ 0.6),以上述实验中的两个比例作为实验条件,对比实验结果如表4所示.

由于TrAdaBoost采用正负样本相同的权重更新策略,其微平均得分差异不大,两个比例下准确率均在90 %以下,虽然其宏平均也达到了0.6,但是对于多类比绝对不平衡问题明显力不从心. 本文算法以类先验概率逼近条件概率的方式建立权重矩阵,能够在

表 4　对比结果

本文算法	宏平均	0.990 1	0.772
	微平均	0.994 4	0.942
TrAdaBoost	宏平均	0.7003	0.608
	微平均	0.8754	0.885

保证多数类样本的情况下兼顾少数类样本,其中宏平均远高于0.6,微平均也在0.94以上,对于训练样本与测试样本不平衡、比例不相同的极端条件依旧可以获得较好的识别效果.

4　结　论

运输系统一旦遭受蓄意攻击,应急救援时需要正确地识别出每一类危险源.然而,少数类样本容易淹没在多数类样本中不易准确识别,如果错误地将少数类危险源识别成其他类危险源,则会使得应急部门响应的应急救援方案出现偏差,不能给予及时的响应,同时会错误地投入大量的人、财、物,使得应急救援的效率低下. 针对蓄意攻击数据的样本有限不均衡的问题,本文提出了多分类器集成加权均衡分布适配的关键危险源识别方法. 在充分选择少数类样本的前提下随机抽样多数类样本,让少数类样本得到充分的训练. 然后构建多分类器集成加权均衡分布适配网络,经过多次迭代得到可靠的目标域伪标签. 利用类先验概率逼近条件分布概率,采用多分类器集成的策略,在提高少数类样本关注度的同时兼顾了多数类样本的识别率. 通过实验和结果分析可知,在解决样本有限不均衡的问题时,较其他几类算法,本文所提出方法不仅保证了整体精度,而且在有效识别少数类样本方面表现更好,更精确地预判出每一类关键危险源为蓄意攻击事件预警和应急救援提供数据支撑. 同时,本文所提出方法也适用于存在样本有限不均衡下的危险源识别问题.

由于数据的可得性,本文仅将不同的攻击方式视为不同的关键危险源,但同一种攻击方式可采用的武器和手段还存在差异,如果能将关键危险源进一步细分,同时在蓄意攻击发生后对其智能识别,则针对性的应急救援工作将更加精准.

参考文献(References)

[1] 李锐, 黄敏. 蓄意攻击下第三方物流可靠性逆向网络设计[J]. 计算机集成制造系统, 2017, 23(9): 1992-2002.
(Li R, Huang M. Reliable reverse network design of third-party logistics under proactive attacks[J]. Computer Integrated Manufacturing Systems, 2017, 23(9): 1992-2002.)

[2] Koivisto R, Kulmala I, Gotcheva N. Weak signals and damage scenarios — Systematics to identify weak signals and their sources related to mass transport attacks[J]. Technological Forecasting and Social Change, 2016, 104: 180-190.

[3] 王立夫, 赵云康, 段乐, 等. 割点失效对复杂网络可控性的影响[J]. 控制与决策, 2019, 34(11): 2310-2316.
(Wang L F, Zhao Y K, Duan L, et al. Effect of cut vertexes-removal on controllability of complex networks[J]. Control and Decision, 2019, 34(11): 2310-2316.)

[4] 李成兵, 张帅, 杨志成, 等. 蓄意攻击下城市群客运交通网络级联抗毁性仿真[J]. 交通运输系统工程与信息, 2019, 19(2): 14-21.
(Li C B, Zhang S, Yang Z C, et al. Invulnerability simulation in urban agglomeration passenger traffic network under targeted attacks[J]. Journal of Transportation Systems Engineering and Information Technology, 2019, 19(2): 14-21.)

[5] 吴迪, 王诺, 于安琪, 等. “丝路”海运网络的脆弱性及风险控制研究[J]. 地理学报, 2018, 73(6): 1133-1148.
(Wu D, Wang N, Yu A Q, et al. Vulnerability and risk management in the Maritime Silk Road container shipping network[J]. Acta Geographica Sinica, 2018, 73(6): 1133-1148.)

[6] National Consortium for the study of terrorism and responses to terrorism (START). Global terrorism database[EB/OL]. (2019-07-30)[2020-08-04]. https: // www.start.umd.edu/research.

[7] O’neil C, Robinson A M, Ingleton S. Mitigating the effects of firebomb and blast attacks on metro systems[J]. Procedia-Social and Behavioral Sciences, 2012, 48: 3518-3527.

[8] Wang J, Jiang C, Yu H. Discrete-event simulation engineering in evaluation of medical treatment capability against biochemical terrorist attacks[J]. Systems Engineering Procedia, 2012, 5: 266-275.

[9] Berman O, Gavious A, Menezes M B C. Optimal response against bioterror attack on airport terminal[J]. European Journal of Operational Research, 2012, 219(2): 415-424.

[10] Nizamani S, Memon N. Detecting terrorism incidence type from news summary[Z]. Berlin, Heidelberg: Springer, 2012, 126: 95-102.

[11] Sivaman R, Srinivasan S, Chandrasekeran R. Big data on terrorist attacks: An analysis using the ensemble classifier approach[EB/OL]. (2015-04-02)[2020-08-04]. https://edlib.net/2015/icidret/icidret2015-

042.pdf.

[12] 肖圣龙, 陈昕, 李卓. 面向社会安全事件的分布式神经网络攻击行为分类方法[J]. 计算机应用, 2017, 37(10): 2794-2798.
(Xiao S L, Chen X, Li Z. Distributed neural network for classification of attack behavior to social security events[J]. Journal of Computer Applications, 2017, 37(10): 2794-2798.)

[13] Meng X, Nie L, Song J. Big data-based prediction of terrorist attacks[J]. Computers & Electrical Engineering, 2019, 77: 120-127.

[14] 李艳霞, 柴毅, 胡友强, 等. 不平衡数据分类方法综述[J]. 控制与决策, 2019, 34(4): 673-688.
(Li Y X, Cai Y, Hu Y Q, et al. Review of imbalanced data classification methods[J]. Control and Decision, 2019, 34(4): 673-688.)

[15] 陈仁祥, 吴昊年, 杨黎霞, 等. 多分类器集成加权均衡分布适配的滚动轴承寿命阶段识别[J]. 仪器仪表学报, 2019, 40(10): 66-73.
(Chen R X, Wu H N, Yang L X, et al. Rolling bearing life stage recognition based on multi-classifier integration of the weighted and balanced distribution adaptation[J]. Chinese Journal of Scientific Instrument, 2019, 40(10): 66-73.)

[16] 么素素, 王宝亮, 侯永宏. 绝对不平衡样本分类的集成迁移学习算法[J]. 计算机科学与探索, 2018, 12(7): 1145-1153.
(Yao S S, Wang B L, Hou Y H. Ensemble transfer learning algorithm for absolute imbalanced data classification[J]. Journal of Frontiers of Computer Science and Technology, 2018, 12(7): 1145-1153.)

[17] 陈琼, 徐洋洋, 陈林清. 不平衡数据的迁移学习分类算法[J]. 华南理工大学学报: 自然科学版, 2018, 46(1): 122-130.
(Chen Q, Xu Y Y, Chen L Q. Transfer learning for classification on unbalanced data[J]. Journal of South China University of Technology: Natural Science Edition, 2018, 46(1): 122-130.)

[18] Wang J, Chen Y, Hao S, et al. Balanced distribution adaptation for transfer learning[C]. IEEE International Conference on Data Mining. New Orleans, 2017: 1129-1134.

[19] Pan S J, Tsang I W, Kwok J T, et al. Domain adaptation via transfer component analysis[J]. IEEE Transactions on Neural Networks, 2011, 22(2): 199-210.

[20] 李慧, 张南南, 曹卓, 等. 基于机器学习的恐怖分子预测算法[J]. 计算机工程, 2020, 46(2): 315-320.
(Li H, Zhang N N, Cao Z, et al. Terrorist prediction algorithm based on machine learning[J]. Computer Engineering, 2020, 46(2): 315-320.)

作者简介

杨黎霞 (1985−), 女, 讲师, 博士生, 从事运输风险管理、大数据分析的研究, E-mail: lixiayang1207@126.com;

许茂增 (1960−), 男, 教授, 博士生导师, 从事商务决策、物流与供应链管理等研究, E-mail: xmzzrxhy@cqjtu.edu.cn;

陈仁祥 (1983−), 男, 教授, 博士生导师, 从事故障诊断、大数据分析等研究, E-mail: manlou.yue@126.com;

吴昊年 (1993−), 男, 博士生, 从事机械设备安全服役、迁移学习的研究, E-mail: 296018167@qq.com.

(责任编辑：孙艺红)