

李夕海, 刘 刚, 刘代志等. 基于最近邻支撑向量特征线融合算法的核爆地震识别. 地球物理学报, 2009, 52(7): 1816~1824, DOI:10.3969/j.issn.0001-5733.2009.07.016

Li X H, Liu G, Liu D Z, et al. Discrimination of nuclear explosions and earthquakes using the nearest support vector feature line fusion classification algorithm. *Chinese J. Geophys.* (in Chinese), 2009, 52(7): 1816~1824, DOI:10.3969/j.issn.0001-5733.2009.07.016

基于最近邻支撑向量特征线融合 算法的核爆地震识别

李夕海, 刘 刚, 刘代志, 秦庆强

第二炮兵工程学院, 西安 710025

摘 要 为增强核爆地震模式分类器的泛化能力以提高对核爆炸事件的准确识别能力, 论文提出了一种选择支撑向量样本集来表征训练样本集的最近邻支撑向量特征线分类算法, 用以训练时扩展核爆地震的训练样本库, 提高分类器的泛化能力. 该算法用于核爆炸和地震的识别结果发现, 和最近邻特征线分类器相比, 提出的算法降低了计算复杂度, 但识别能力却有些许降低. 对新算法的分析发现, 纯粹的支撑向量集不能完全代表原始样本空间集, 支撑向量比例在其中具有重要作用, 为发挥支撑向量比例的作用以提高核爆分类器的识别能力, 提出了最近邻支撑向量特征线融合算法. 最后以核爆地震数据库对上述算法进行了检验和分析, 理论分析和识别结果证实, 在相同的训练样本选择条件下, 最近邻支撑向量特征线融合算法对于核爆炸的识别来说具有较好的泛化能力, 正确识别率达到 90.3%, 且优于支持向量机算法和最近邻特征线算法.

关键词 最近邻特征线, 支撑向量, 核爆炸, 天然地震

DOI:10.3969/j.issn.0001-5733.2009.07.016

中图分类号 P315

收稿日期 2007-11-13, 2009-06-09 收修定稿

Discrimination of nuclear explosions and earthquakes using the nearest support vector feature line fusion classification algorithm

LI Xi-Hai, LIU Gang, LIU Dai-Zhi, QIN Qing-Qiang

Xi'an Research Institute of High Technology, Xi'an 710025, China

Abstract Statistical pattern recognition is an important method to distinguish nuclear explosions from natural earthquakes, but because the number of nuclear explosions is few and this limits the generalization performance of the nuclear explosions' classifier, and finally affects the classification performance for the nuclear explosions. To improve the classification performance of the classifier, a nearest support vector feature line algorithm is introduced to expand the virtual nuclear explosions samples and improve the generalization performance correspondingly. However, this new algorithm can not distinguish nuclear explosions from natural earthquakes more accurately, and even is not as good as the nearest feature line algorithm. Theoretical analysis finds that the support vector set is not a good representation of the train set and the ratio of support vector plays an important role to the classifier. To improve the recognition result with the role of support vector ratio, support vector machine is integrated with nearest support vector

feature line method by support vector ratio, and then a nearest support vector feature line fusion classification algorithm is presented, and experimental results and corresponding analysis show that this novel algorithm is more appropriate to the classification of underground nuclear explosions and natural earthquakes, and the accurate classification ratio is 90.3% on the data set of nuclear explosions and natural earthquakes, and moreover this classification ratio is higher than support vector machine and nearest feature line.

Keywords Nearest feature line, Support vector, Nuclear explosion, Natural earthquake

1 引言

1996年,全面禁止核试验条约(comprehensive test banning treaty, CTBT)在日内瓦签订,为监测各国是否遵守禁核条约,将全球地震波监测方法列在首位,这使地下核爆炸和天然地震的分类问题成为关键问题.目前对地下核爆炸事件识别的研究,主要从两个方向进行.一是从地球物理学的角度,主要是地震学的角度,通过分析震源机制、地震波走时、各种震相,计算各种分类识别参数(也即识别特征),如各种震相幅度比、震级比等等,进而达到分类识别之目的.二是从信号与信息处理的角度,通过对地震波形所构成的时间序列进行各种分析(变换),提取对分类识别有用的各种数值特征,如波形复杂度、谱比值、各种矩、实倒谱、时频谱、分形维等等,进而设计各种分类识别器进行分类识别,即核爆地震模式识别.在核爆地震模式识别领域,从统计模式识别的角度出发来研究核爆地震识别问题一直是军控核查工作努力的重要方向,很多学者提出了多种方法进行核爆地震的分类识别,例如杨选辉、孙煜等^[1,2]分别提出了应用小波包变换方法,沈萍等^[3]提出了应用瞬态谱方法,何永锋等^[4]提出应用经验格林函数方法,刘代志等^[5]提出了应用分形分析方法识别地下核爆炸.此外,刘代志等^[6~14]提出了一系列识别方法.但由于地下核爆炸试验的次数是有限的,因此,上述大多数文献是基于有限的核爆炸事件基础上做出的识别结论,有的甚至是基于一个或几个核爆炸事件得出的结论,因此提出的各种判据和分类方法对未知事件的识别能力也不能完全确定.这是因为从统计模式识别的角度出发进行地下核爆炸和天然地震的自动识别,必须要求应用已知事件性质的样本(即训练样本)对分类器的参数进行训练,因此从理论上说只有训练样本足够多才能真正反映两类事件的真实分布,这样获得的分类器才能对未知事件的分类能力最强.因此,从统计模式识别的角度

研究核爆地震识别问题必然希望训练样本越多越好.那么,在现有地下核爆炸事件十分有限的条件下,能否扩展核爆炸样本,以使获得的核爆地震分类器对未知地震事件推广识别性能更好呢,即具有更好的识别能力?论文针对这个问题在最近邻特征线算法的基础上展开了研究,提出了一种最近邻支撑向量特征线融合分类算法,使它用于地下核爆炸和天然地震的识别时,既具有较好的泛化能力,同时具有较好的分类识别能力.

2 最近邻特征线算法及其分析

最近邻特征线(Nearest Feature Line)算法是 Stan Z. Li 和 Lu Juwei^[15]在 1999 年提出的一种用于人脸识别的分类算法.该方法是在最近邻分类方法的基础上提出的,它充分利用了训练样本空间中任意两个特征的连线(即特征线)来扩展属于同一类别的虚拟样本,即逼近了两个样本的一些变化,如人脸姿态、表情变化和光照条件的变化,而分类则通过待分类样本与特征线之间的距离来进行,因此该方法无形中使某一类的样本能得到无穷多的扩展.该方法提出后,被应用到了人脸识别^[16]、雷达目标识别^[17]及汽车牌照识别^[18]等领域,并且文献^[19]对该方法进行了改进以适用于视频数据的内容检索.

最近邻特征线分类的前提是对于待分类的每个样本至少有两个特征(此条件绝大多数情况下都能满足),它通过特征空间中的两个样本点之间连成直线(即形成特征线)来完成线性内插或外推.对于通过同一类的两个样本点 X_1 和 X_2 的直线记作 $\overline{X_1 X_2}$,并称它为该类的一条特征线,待分类样本点 X 在特征线 $\overline{X_1 X_2}$ 上的投影点为 p (见图 1),则样本点 X 与特征线 $\overline{X_1 X_2}$ 的距离可以定义为

$$d(X, \overline{X_1 X_2}) = \|X - p\|, \quad (1)$$

其中 $\|\cdot\|$ 表示范数.

投影点 p 可表示为 $p = X_1 + u(X_2 - X_1)$, 其中

$u \in \mathbf{R}$, 可称为位置参数, u 的值可由 X, X_1 和 X_2 通过计算得到. 由于直线 \overline{pX} 垂直于特征线 $\overline{X_1X_2}$, 因此 $\langle (p - X) \cdot (X_2 - X_1) \rangle = \langle [X_1 + u(X_2 - X_1) - X] \cdot (X_2 - X_1) \rangle = 0$, (2)

所以得到

$$u = \frac{\langle (X - X_1) \cdot (X_2 - X_1) \rangle}{\langle (X_2 - X_1) \cdot (X_2 - X_1) \rangle}, \quad (3)$$

其中, $\langle \cdot \rangle$ 表示两个矢量的内积, 参数 u 描述了投影点 p 与样本点 X_1 和 X_2 的联系. 当 $u=0$ 时, $p=X_1$; 当 $u=1$ 时, $p=X_2$; 当 $0 < u < 1$, p 为在样本点 X_1 和 X_2 之间的一个内插点; 当 $u > 1$ 时, p 在特征线 $\overline{X_1X_2}$ 上, 且在点 X_2 外; 当 $u < 0$ 时, p 在特征线 $\overline{X_1X_2}$ 上, 且在点 X_1 外.

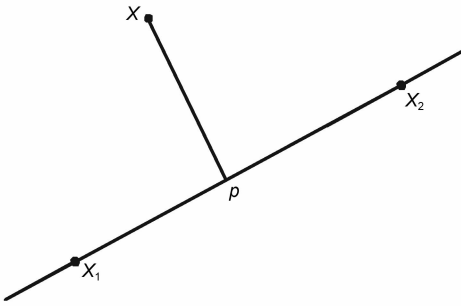


图 1 通过两个样本点 X_1 和 X_2 构成的特征线 $\overline{X_1X_2}$

Fig. 1 Feature Line $\overline{X_1X_2}$ based on two samples X_1 and X_2

特征线为两个样本的线性变化提供了信息, 实际上它为某一类样本集提供了无数多个虚拟的特征点, 因此是对训练样本特征集的一种无形扩充. 假定共有 M 类样本, 其中第 c 类有 N_c 个特征点 ($c=1, 2, \dots, M$), 则待分类样本点 X 与第 c 类的两个样本点 X_i^c 和 X_j^c ($i \neq j$) 组成的特征线 $\overline{X_i^cX_j^c}$ 之间的距离为 $d(X, \overline{X_i^cX_j^c})$. X 与第 c 类每条特征线构成的距离个数总共是 $N_{c_{total}} = N_c(N_c - 1)/2$, 则 X 与 M 类共存 $N_{total} = \sum_{c=1}^M N_{c_{total}}$ 个距离度量. 将所有的距离按升序排列, 则构成最小距离的特征线(用 $\overline{X_i^{c^*}X_j^{c^*}}$ 表示)所在的类别即为 X 的类别, 即

$$d(X, \overline{X_i^{c^*}X_j^{c^*}}) = \min_{1 \leq c \leq M} \min_{1 \leq i < j \leq N_c} d(X, \overline{X_i^cX_j^c}). \quad (4)$$

通过以上阐述, 可以看出, 最近邻特征线分类算法的优点就在于通过待识别样本对最近邻距离的求解无形中扩展了训练样本集. 但该算法带来的问题是, 若训练样本数量较多时, 如第 c 类的样本容量为 N_c , 则对一个未知样本的分类至少需要进行 $\sum_{c=1}^M C_{N_c}^2$

次距离(内积)运算(m 为类别数目), 即使 $N_c < 100$, 计算复杂度也十分可观. 而在很多实际问题中, 训练样本的个数都大于 100, 因此应用最近邻特征线分类算法时必须考虑计算复杂度的问题.

3 最近邻支撑向量特征线分类算法及应用

3.1 算法阐述

最近邻特征线算法充分利用训练样本集中的特征线来扩展属于同一类别的虚拟样本, 而分类则通过待分类样本与特征线之间的距离来进行, 因此无形中使训练样本得到无穷多的扩展, 训练样本的增多必然提高了分类识别的泛化能力, 这无疑有益于分类器的设计. 然而, 如前所述, 虽然对传统的统计模式识别方法而言, 训练样本是有限的, 但对特征线方法来说, 由于扩展了有限的训练样本, 其计算复杂度在很多情况下也是十分可观的. 能否既能提高分类器的泛化能力, 而其计算复杂度又不随之迅速增加呢? 我们知道, 支撑向量机^[20~26]是一种建立在统计学习理论和结构风险最小化原理基础上的学习机器, 它根据有限样本信息在模型的复杂度(即对特定训练样本的分类精度)与学习能力(即无错误识别任意样本的能力)之间寻求最佳折衷, 以获得最强的泛化能力, 是一种针对小样本集的分类识别算法. 支撑向量机的特点是在训练样本集中寻找一个支撑向量集, 通过对支撑向量集中样本的线性加权来获得权矢量 $W = \sum_{i \in SV} y_i \alpha_i X_i$ (SV 表示训练样本集中的支撑向量集, X_i 为支撑向量集中的第 i 支撑向量样本, α_i 为 X_i 对应的权系数)进而对未知样本进行分类. 也即是说, 支撑向量机分类器的设计是由训练样本集中的支撑向量样本确定, 支撑向量样本对于分类具有决定性影响, 而非支撑向量样本不参与支撑向量机分类器的设计中, 即它对于未知样本的分类不起作用. 借助支撑向量机设计的思想, 本文提出通过选择训练样本集中的支撑向量来代替整个训练样本集实现最近邻特征线分类, 即最近邻支撑向量特征线分类算法. 研究该算法能否在降低计算复杂度的同时, 又能保持最近邻特征线分类算法的识别性能. 算法具体概述如下:

①对训练样本集计算其支撑向量集, 即找出权值系数 $\alpha_i > 0$ 对应的训练样本构成支撑向量集(关于如何寻找训练样本集中的支撑向量集, 可参考文

献[20~26]).

②将支撑向量按训练样本集中对应的相应类别进行分类,形成各类对应的支撑向量训练样本集.

③用支撑向量训练样本集代替整个训练样本集,在支撑向量集上实现最近邻特征线的选取和对测试样本的分类.

3.2 分类识别实验及结果分析

将最近邻支撑向量分类算法,分别以线性支撑向量和非线性支撑向量两种情况进行了地下核爆炸与天然地震的分类实验(选择径向基核函数,即 $k(x, y) = \exp(-\|x - y\|^2 / (\sigma^2))$,其中 x, y 表示参与核函数运算的两个样本的特征向量, σ 为待定参数,一般情况下通过训练获得其参数值).参与分类识别的核爆地震数据集共有短周期国外核爆炸和天然地震事件波形各 100 个(采样频率 40 Hz),核

爆炸和天然地震数据采集的地震台站及每个台站采集的事件数据个数见表 1. 每个事件采用的是其地震台站采集的垂直分量波形,其波形见图 2.

对于最近邻支撑向量算法的分类识别实验,训练样本集为从核爆地震数据集中随机选取核爆炸和地震事件波形各 50 个,其余样本构成测试样本集.对于每个事件波形信号提取 12 类特征量(分别为:一阶相关(1)、一阶谱相关(2)、二阶相关(3)、二阶谱相关(4)、三阶相关(5)、三阶谱相关(6)、波形复杂度(7)、短时谱(8)、谱比值(9)、频率三次矩(10)、AR 模型系数(11)和实倒谱特征(12),具体提取方法参考文献[27].由于各类特征的计算方法不同,其值大小差异很大,因此在进行分类实验前分别对各类特征进行了归一化处理,同时为了抵消随机选取训练样本和测试样本给分类识别结果带来的影响,对该

表 1 本文使用的中国地震台站坐标和各台站采集的事件数

Table 1 Coordinates of earthquake observatories in China and numbers of events

编号	台站	代号	北纬(°)	东经(°)	海拔(m)	天然地震事件个数	地下核试验事件个数
1	北京	BJI	40.0403	116.1750	43	13	16
2	兰州	LZH	36.0867	103.8444	1560	15	7
3	恩施	ENH	30.2717	109.4868	487	12	15
4	昆明	KMI	25.1232	102.7400	1945	8	10
5	琼中	QIZ	19.0293	109.8432	230	3	18
6	上海	SSE	31.0956	121.1867	10	12	11
7	乌鲁木齐	WMQ	43.8211	87.6950	970	18	6
8	海拉尔	HIA	49.2666	119.1666	610	11	8
9	牡丹江	MDJ	44.6163	129.5918	250	8	9

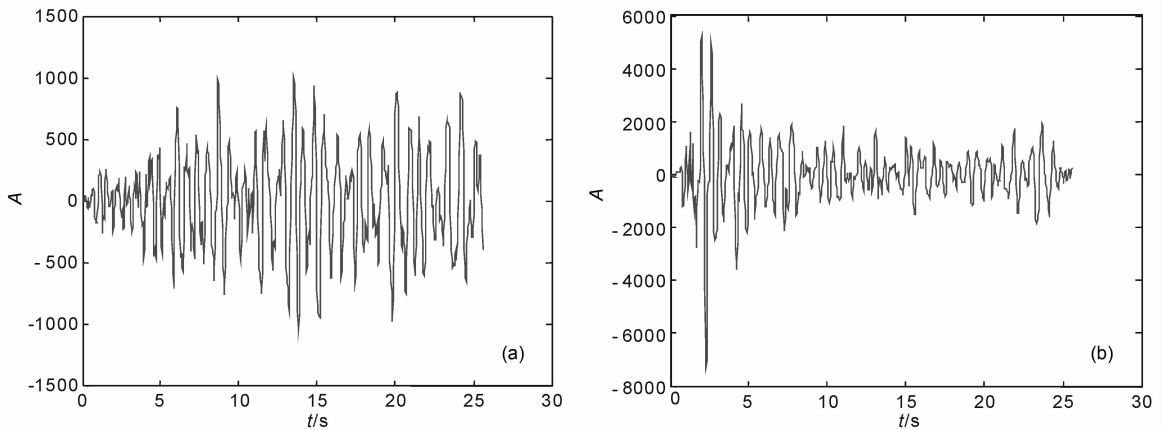


图 2 恩施台站监测到的地震波形记录

(a) 1987-02-17 监测的一次天然地震波形;(b)1987-04-03 监测到的某次国外核爆炸地震波形记录.

Fig. 2 (a) A natural earthquake waveform observed by ENH earthquake observatory on Feb. 17 1987

(b) A nuclear explosion waveform observed by ENH earthquake observatory on Apr. 3 1987

分类实验重复进行了 20 次, 实验结果为取 20 次分类结果的平均. 为便于和支撑向量机及最近邻特征线算法进行比较, 同时应用该数据库进行了支撑向量机和最近邻特征线算法的分类实验. 线性支撑向量条件下的分类识别结果见表 2, 非线性支撑向量条件下的分类识别结果见表 3, 三种算法得到的核爆地震分类错误率曲线见图 3.

对表 2 与表 3 以及图 3 进行对比分析, 可得到如下几点结论:

(1) 若支撑向量比例较小, 则支撑向量机算法获得的核爆地震分类识别结果较好, 且通过选择合适的参数 σ , 非线性支撑向量机算法获得的最小错误率达到了 11.25%, 优于线性支撑向量机;

(2) 由于每次分类实验对训练样本集的选择是

随机的, 导致了表 3 中最近邻特征线算法的平均识别结果有些差别, 但变化相对较小;

(3) 对最近邻支撑向量特征线分类算法, 其分类结果和其他两种算法相比, 识别能力并不好, 绝大多

表 2 线性支撑向量条件下的分类结果

Table 2 Classification results on the condition of linear support vectors

分类算法	分类错误率 (%)	错误率方差	支撑向量比例 (%)
支撑向量机 (LSVM)	11.30	8.5368	平均 29
基于支撑向量的 NFL	31.55	69.4148	最大 39
最近邻特征线 (NFL)	12.05	11.6289	最小 21

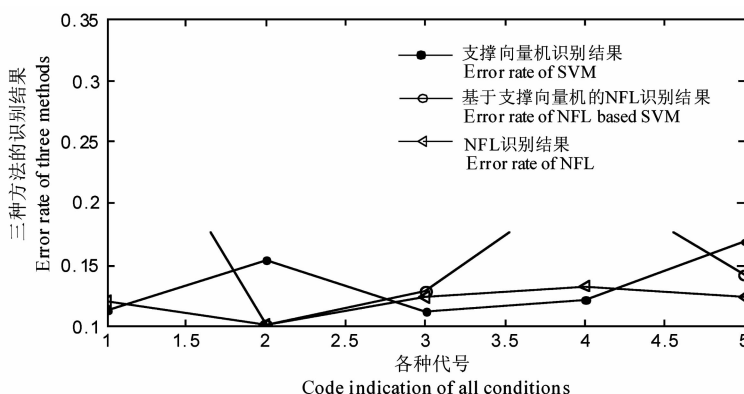


图 3 三种算法得到的核爆地震分类错误率曲线

其中横坐标 1 代表线性支撑向量条件, 横坐标 2, 3, 4 分别代表非线性支撑向量条件下参数 σ 从小到大的不同取值.

Fig. 3 Error rate curves of nuclear explosions and earthquakes for three recognition algorithms

"1" on the x -axes denotes the condition of linear support vector, and "2, 3, 4" denote the various condition of nonlinear support vector with different parameter σ .

数情况下甚至比最近邻特征线算法的识别结果还差;

(4) 支撑向量的比例越小, 最近邻支撑向量特征线算法的分类结果越差; 支撑向量比例增大后, 分类效果会有所改善, 当支撑向量比例达到 100% 时, 其分类结果和最近邻特征线算法结果一致.

(5) 在非线性支撑向量条件下, 径向基核函数中参数 σ 的选择对于核爆炸分类器的识别能力起重要作用, 不同的参数导致的分类识别结果差别较大 (目前对分类器中参数 σ 的选择一般是参数空间中搜索选择其最优值, 目前还没有确定的参数选择方法).

最近邻支撑向量特征线算法获得的分类效果较差, 对其算法分析后发现, 支撑向量比例是针对支撑向量机提出的, 它是对经核函数映射后的特征空间类别可分性的一种有效度量, 如果支撑向量比例较

小, 则经核函数映射后的特征空间中用于区分两类的支撑向量样本数目较少, 表明映射后的特征空间中两类样本的可分性较好; 若支撑向量比例较大, 则经核函数映射后的特征空间中用于区分两类的支撑向量样本数目就比较多, 表明映射后的特征空间中两类样本的可分性不太好. 支撑向量机算法对样本的分类是在经核函数映射后的特征空间进行, 即支撑向量机通过权值参数 α_i 定量地给出了每个支撑向量样本在泛化推广时的重要性 (权值), 而最近邻支撑向量特征线分类算法虽然选择了支撑向量样本作为训练样本, 但其分类运算确是在原始特征空间中进行, 也就是说, 由于最近邻支撑向量特征线分类算法仅应用经核函数映射后的特征空间选择了支撑向量样本, 而没有在相应的特征空间中进行分类, 却

表 3 非线性支撑向量条件下的分类结果
(径向基核函数)

Table 3 Classification results on the condition of
nonlinear support vectors (RBF Kernel Function)

参数 σ	分类算法	分类错误率 (%)	错误率方差	支撑向量比例 (%)
0.1	支撑向量机 (SVM)	15.4	9.6	平均 100
	基于支撑向量的 NFL	10.1	4.9889	最大 100
	最近邻特征线(NFL)	10.1	4.9889	最小 100
1	支撑向量机 (SVM)	11.25	3.4605	平均 75.85
	基于支撑向量的 NFL	12.9	8.2	最大 82.00
	最近邻特征线(NFL)	12.35	7.7132	最小 70.00
5	支撑向量机 (SVM)	12.2	3.9556	平均 52.5
	基于支撑向量的 NFL	22	36	最大 60.0
	最近邻特征线(NFL)	13.3	3.5667	最小 48.0
10	支撑向量机 (SVM)	16.9	4.1	平均 74.0
	基于支撑向量的 NFL	14.2	22.8444	最大 78.0
	最近邻特征线(NFL)	12.4	17.6	最小 71.0

选择在原始空间中进行分类. 在原始空间中, 由于支撑向量样本不能准确表征其空间中类别的可分性, 因此, 分类识别结果必然不会很好. 但随着支撑向量比例的增大, 可应用的特征点(训练样本)越来越多, 当支撑向量比例达到 100% 时, 已应用所有的特征点进行分类, 实质已是最近邻特征线算法, 所以其分类效果与最近邻特征线算法的识别结果一致.

通过以上的比较与分析, 可以看出, 最近邻支撑向量特征线算法不能完全代表最近邻算法, 虽然降低了一定的计算复杂度, 但对识别结果的影响也比较大, 因此必须根据支撑向量样本和相应的支撑向量比例对该算法进一步完善.

4 基于最近邻支撑向量特征线融合的分类器设计及应用

4.1 算法阐述

从对最近邻支撑向量分类算法的分析可以看出, 在对核爆地震事件进行分类时, 若支撑向量比例较小时, 说明此时经核函数映射后的特征空间中核

爆炸和天然地震之间的类间可分性较好, 在大多数情况下支撑向量机算法获得的分类结果较好, 但此时由于最近邻支撑向量特征线算法仅用支撑向量样本, 导致分类效果较差; 与之相反, 若支撑向量比例较大时, 说明此时经核函数映射后的特征空间中核爆炸和天然地震之间的类间可分性不太好, 大多数情况下会导致支撑向量机的分类误差增大, 而此时最近邻支撑向量特征线算法由于支撑向量的增多会产生较好的分类结果. 针对这两种仅适用不同类别可分性情况的分类算法存在的缺陷, 能否将两种方法融合起来, 使类别样本可分性在两种情况下都能取得较好的分类结果呢? 由于支撑向量比例对支撑向量机和最近邻支撑向量特征线算法具有直接影响, 并且与支撑向量机的分类正确率大体成反比关系, 而与最近邻特征线算法大体成正比关系, 因此, 将支撑向量比例作为一种加权系数对两种方法进行融合, 称之为基于最近邻支撑向量特征线融合(NSVFLF)的分类算法. 该算法(以两类样本的分类为例)的具体步骤如下:

①对训练样本集计算其支撑向量样本, 即找出权值系数 $\alpha_i > 0$ 对应的训练样本构成支撑向量集, 并计算支撑向量比例 η .

②将支撑向量按训练样本集的相应类别进行归类, 形成各自新的支撑向量训练样本集. 即第一类训练样本集 C_1 中的支撑向量集为 $SV1$, 第二类训练样本集 C_2 中的支撑向量集为 $SV2$.

③将每一类中的支撑向量作为特征点, 对待识别样本计算其到每一类中任意两点构成的特征线的距离, 并找出其到每一类的最小距离. 即对于待识别样本 z , 分别计算 z 到两类训练样本集 C_1 和 C_2 对应支撑向量集 $SV1$ 和 $SV2$ 的所有特征线的距离, 并找到与 $SV1$ 和 $SV2$ 的最小距离, 分别设为 d_1^z 和 d_2^z . 若 $d_2^z - d_1^z > 0$, 则 $z \in C_1$, 反之, 若 $d_2^z - d_1^z < 0$, 则 $z \in C_2$.

④对于待识别样本 z , 按照支撑向量机判别函数 $f(z, \alpha, b) = \sum_{i \in SV} y_i \alpha_i \langle X_i \cdot z \rangle + b$ 计算其在支撑向量机中的分类结果(其中 X_i 为第 i 个支撑向量样本, α_i 为 X_i 对应的权系数, $SV = SV1 \cup SV2$, $y_i \in \{+1, -1\}$ 为样本 X_i 属于的类别号, b 为通过一个支撑向量求解的待解参数). 即, 若 $f(z, \alpha, b) > 0$, 则 $z \in C_1$; 反之, 则 $z \in C_2$.

⑤将第③步计算的最近邻特征线分类结果和第④步计算的支撑向量机分类结果通过支撑向量比例

按下式进行融合,进行最终的属性判别:

$$C(z) = \eta(d_2^z - d_1^z) + (1 - \eta)f(z, \alpha, b). \quad (5)$$

若 $C(z) > 0$, 则 $z \in C_1$; 反之, 若 $C(z) < 0$, 则 $z \in C_2$.

上述算法中, 若第一类训练样本集 C_1 的样本数为 N_1 , 第二类训练样本集 C_2 的样本数为 N_2 , 训练样本集中总的支撑向量样本个数为 m , 则算法流程见图 4.

4.2 分类识别实验及结果分析

为验证 NSVFLF 算法的有效性, 在与表 1、2 同样的训练样本选择条件下(每次计算时随机选取的样本一致, 获得的支撑向量一致), 在核爆地震数据集中进行了核爆炸和天然地震的分类识别, 分类结果见表 4.

表 4 NSVFL 的分类结果

Table 4 Classification results of NSVFL

分类算法	分类错误率(%)	错误率方差
NSVFLF(线性条件)	11.25	6.0921
NSVFLF($\sigma=0.1$)	10.1	4.9889
NSVFLF($\sigma=1$)	9.7	3.1684
NSVFLF($\sigma=5$)	11.9	2.7667
NSVFLF($\sigma=10$)	12.1	6.9889

将表 4 中线性支撑向量条件下得到的分类结果与表 2 对比分析, 发现此时支撑向量比例较小, 说明支撑向量机算法中特征空间的类别可分性较好, 因此支撑向量机算法获得的识别结果优于最近邻特征线算法, 而 NSVFLF 算法获得的错误识别率比支撑

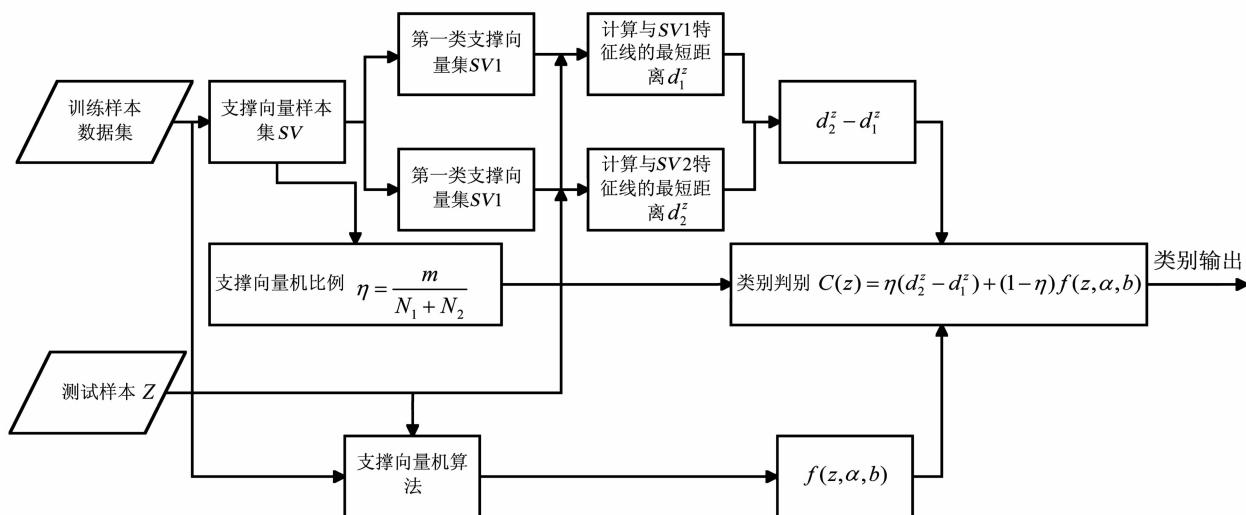


图 4 最近邻支撑向量特征线融合分类算法流程图

Fig. 4 Flow chart of nearest support vector feature line fusion algorithm

向量机算法还低 0.5%.

将表 4 中在非线性支撑向量条件下的识别结果与表 3 对比分析, 可得到如下结论:

(1) 参数 $\sigma=0.1$ 时, 支撑向量比例达到 100%, 说明此时支撑向量机算法获得的特征空间类别可分性十分不好, NSVFLF 算法公式 (5) 中 $\eta=1$, NSVFLF 算法变为最近邻特征线算法, 因此它与最近邻特征线算法获得的分类结果一致, 并且优于支撑向量机算法的识别结果.

(2) 参数 σ 选取 0.1, 1, 5, 10 四种取值情况时, 支撑向量比例由大变小, 再变大, 说明经核函数变换后的特征空间的类别可分性参数值应该在 [0.1, 10] 这个区间内, 因此可通过区间搜索优化的方法获得这个最佳值.

(3) 对表 4 中 NSVFLF 的识别结果与对应的支撑向量比例之间的对比关系进行分析, 发现在现有的识别数据下, 并不是支撑向量的比例越小, 获得的分类识别结果越好.

(4) 无论是线性支撑向量条件下, 还是非线性支撑向量条件下, NSVFLF 算法获得的错误率方差比其他三种算法的要小, 说明 NSVFLF 算法对训练样本集的随机选取具有的较好不敏感性, 优于其他三种方法, 因此 NSVFLF 算法应对训练样本集的要求具有更宽松的约束.

总之, 从表 4 中的识别结果可以发现, 即使随机选取的参数 σ 不是最优值的条件下, 而 NSVFLF 算法获得的分类识别结果都优于或等于表 3 中三种算法获得的最佳分类结果, 并且具有较小的错误率方差.

5 结 语

为提高核爆地震自动识别中分类器的泛化能力并最终提高核爆炸的正确识别能力,在将支撑向量引入到最近邻特征线分类算法的基础上,提出了一种最近邻支撑向量特征线融合分类算法,并将其应用到核爆炸和天然地震的分类识别中.在短周期核爆地震数据库上进行的分类实验结果表明,在相同的核函数及对应参数值的条件下,最近邻支撑向量特征线融合分类算法对核爆地震两类事件的识别结果优于专门针对小样本集学习而提出的支撑向量机识别结果.但应当指出的是该算法与支撑向量机算法一样,存在一个寻找合适的核函数以及相应的最佳参数值的过程,现有寻找核函数的方法是在已有的几个核函数中尝试选择一个最佳核函数,而能否针对核爆炸的分类识别问题构造一个适合它的核函数正是我们下一步需要研究的问题.

参考文献 (References)

- [1] 杨选辉,沈 萍,刘希强等.地震与核爆识别的小波包方法.地球物理学报,2005,48(1):148~156
Yang X H, Shen P, Liu X Q, et al. Application of method of spectral component ratio of wavelet-packets to discrimination between earthquakes and nuclear explosions. *Chinese J. Geophys.* (in Chinese), 2005, 48(1):148~156
- [2] 孙 煜,范万春,许 进等.基于小波包变换的地震事件分类.核电子学与探测技术,2005,25(1):32~36
Sun Y, Fan W C, Xu J, et al. Seismic signal classification based on wavelet packet transform. *Nuclear Electronics & Detection Technology* (in Chinese), 2005, 25(1):32~36
- [3] 沈 萍,郑治真.瞬态谱在地震与核爆识别中的应用.地球物理学报,1999,42(2):231~240
Shen P, Zheng Z Z. Application of transient spectrum to discrimination of nuclear explosions and earthquakes. *Chinese J. Geophys.* (in Chinese), 1999, 42(2):231~240
- [4] 何永锋,陈晓非.利用经验格林函数识别地下核爆炸与天然地震.中国科学(D 辑 地球科学),2006,36(2):177~181
He Y F, Chen X F. Using empirical Green' function to discriminate underground nuclear explosions and natural earthquakes. *Science in China (Ser. D Earth Sciences)* (in Chinese), 2006, 36(2):177~181
- [5] Liu D Z, Zhao K, Zou H X, et al. Fractal analysis with applications to seismological pattern recognition of underground nuclear explosions. *Signal Processing*, 2000, 80(9):1849~1861
- [6] 刘代志,张 斌,李夕海等.核爆地震模式识别中的特征相空间研究.地球物理学报,2006,49(3):871~876
Liu D Z, Zhang B, Li X H, et al. On the feature phase space in the seismic pattern recognition of underground nuclear explosions. *Chinese J. Geophys.* (in Chinese), 2006, 49(3):871~876
- [7] 刘代志,李夕海,何志文等.基于非线性加权 K-相关的核爆地震识别.核电子学与探测技术,2005,25(6):587~589
Liu D Z, Li X H, He Z W, et al. Identification of underground nuclear explosions and natural earthquakes based on K-correlation of nonlinear weighting. *Nuclear Electronics & Detection Technology* (in Chinese), 2005, 25(6):587~589
- [8] 张 斌,李夕海,苏 娟等.基于支持向量机的核爆地震自动识别.核电子学与探测技术,2005,25(1):44~47
Zhang B, Li X H, Su J, et al. Recognition of nuclear explosion and natural earthquake based on SVM. *Nuclear Electronics & Detection Technology* (in Chinese), 2005, 25(1):44~47
- [9] 刘代志,邹红星,韦荫康等.分形分析与核爆地震模式识别.模式识别与人工智能,1997,10(2):153~158
Liu D Z, Zou H X, Wei Y K, et al. Fractal analysis with applications to seismic pattern recognition of nuclear explosion. *Pattern Recognition and Artificial Intelligence* (in Chinese), 1997, 10(2):153~158
- [10] 刘代志,邹红星,赵 克等.吸引子分析与核爆地震模式识别.电子学报,1997,25(10):122~125
Liu D Z, Zou H X, Zhao K, et al. Attractor analysis in seismic pattern recognition of nuclear explosion. *Acta Electronica Sinica* (in Chinese), 1997, 25(10):122~125
- [11] Liu D Z, Wang R M, L I X H, et al. Seismic pattern recognition of nuclear explosion based on generalization learning algorithm of BP network and genetic algorithm. *IEEE International Symposium on Neural Networks*, 2004. 959~964
- [12] Li X H, Zhao K, Liu D Z, et al. Feature extraction and identification of underground nuclear explosion and natural earthquake based on FM^mlet transform and BP neural network. *IEEE International Symposium on Neural Networks*, 2004. 925~930
- [13] Liu D Z, Li X H, Zhang B. Feature selection and identification of underground nuclear explosion and natural earthquake based on gamma test and BP neural network. *IEEE International Symposium on Neural Networks*, 2005. 393~398
- [14] Liu G, Li X H, Liu D Z, et al. Feature extraction of underground nuclear explosions based on NMF and KNMF. *IEEE International Symposium on Neural Networks*, 2006. 1400~1405
- [15] Stan Z Li, Lu J W. Face recognition using the nearest feature line method. *IEEE Trans. Neural Networks*, 1999, 10(2):439~443
- [16] 叶伊松,武 研.基于 ICA 和 NFL 分类的局部人脸识别方法.中国图象图形学报,2005,10(4):468~472

- Ye Y S, Wu Y. ICA/NFL local face recognition. *Journal of Image and Graphics* (in Chinese), 2005, **10**(4):468~472
- [17] 孟继成, 杨万麟. 雷达目标识别中的双距离像子空间法. 系统工程与电子技术, 2004, **26**(6):724~725
- Meng J C, Yang W L. Range profile recognition of radar target based on dual range subspace. *Systems Engineering and Electronics* (in Chinese), 2004, **26**(6):724~725
- [18] 谢建平. 一种基于快速最近邻特征线的汽车牌照识别方法. 视频技术应用与工程. 2006, **4**:93~96
- Xie J P. A method of license character recognition based on fast nearest feature line. *Application & Project of Video Technologies* (in Chinese), 2006, **4**:93~96
- [19] 赵黎, 祁卫, 李子青等. 利用改进 NFL 算法对镜头进行基于内容的检索. 软件学报, 2002, **13**(4):586~590
- Zhao L, Qi W, Li Z Q, et al. Content-based retrieval of video shot using the improved nearest feature line method. *Journal of Software* (in Chinese), 2002, **13**(4):586~590
- [20] Nello Cristianini, John Shawe-Taylor. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. England; Press of the University of Cambridge, 2000
- [21] Boser B E, Guyon I M, Vapnik V N. A training algorithm for optimal margin classifiers. In: Haussler D eds. Proceedings of the 5th Annual ACM Workshop on Computational Learning Theory, 1992. 144~152
- [22] Cortes C, Vapnik V N. Support Vector Networks. *Machine Learning*, 1995, **20**: 273~297
- [23] Vapnik V N. The Nature of Statistical Learning Theory. New York; Springer-Verlag, 1995
- [24] Schölkopf B. Support Vector Learning. Munich; Oldenbourg Verlag, 1997
- [25] Vapnik V N. Statistical Learning Theory. Wiley: New York, 1998
- [26] Schölkopf B, Burges C J C, Smola A J. Advances in Kernel Methods—Support Vector Learning. Cambridge: MIT Press, 1999
- [27] 李夕海. 核爆地震特征提取与识别算法研究[博士学位论文]. 西安: 第二炮兵工程学院, 2006
- Li X H. Feature extraction and target recognition of underground nuclear explosions and natural earthquakes [Ph. D. thesis]. Xi'an: The Second Artillery Institute of Engineering, 2006

(本文编辑 汪海英)