0254-6124/2019/39(3)-295–08 *Chin. J. Space Sci.* 空间科学学报

YE Yudong, FENG Xueshang. Study on geoeffectiveness of interplanetary coronal mass ejections by support vector machine (in Chinese). *Chin. J. Space Sci.*, 2019, **39**(3): 295-302. DOI:10.11728/cjss2019.03.295

行星际日冕物质抛射地磁效应研究的支持向量机方法初步研究 *

叶煜东 [1,2] 冯学尚 [1]

1(中国科学院国家空间科学中心 空间天气学国家重点实验室 北京 100190)

2(中国科学院大学 北京 100049)

摘 要 行星际日冕物质抛射 (Interplanetary Coronal Mass Ejection, ICME) 与地球磁层相互作用并带来地磁暴等地磁扰动. 从 Richardson 和 Cane 提供的近地球 ICME 列表中筛选出 ICME 事件集, 基于 ICME 扰动期间的行星际等离子体与磁场数据提取出特征. 通过计算各特征的费舍尔分值 (Fisher Score), 对这些特征进行选择, 发现行星际磁场南北向分量持续时间小于 -10 nT 且激波等扰动所带来的 ICME 扰动开始时, 太阳风速度的增量等特征与 ICME 事件的地磁效应密切相关. 这与现有的传统统计研究结果一致. 以这些特征为基础, 训练得到的径向基函数支持向量机能够以 0.78±0.08 的准确率判断 ICME 事件是否会产生中等及以上强度的地磁暴 ($Dst \leqslant -50$ nT).

关键词 支持向量机, 行星际日冕物质抛射, 地磁效应

中图分类号 P 353

Study on Geoeffectiveness of Interplanetary Coronal Mass Ejections by Support Vector Machine

YE Yudong[1,2] FENG Xueshang[1]

1(*State Key Laboratory of Space Weather, National Space Science Center, Chinese Academy of Sciences, Beijing* 100190)

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

Abstract As arriving at the Earth, Interplanetary Coronal Mass Ejections (ICME) will interact with the Earth's magnetosphere and cause geomagnetic storms. The ICME event set is obtained by Richardson and Cane's Near Earth ICME list, and the input features are extracted based on interplanetary solar wind and magnetic data during ICME disturbance. A total of 483 ICME events from 1996 to 2006 are chosen in this study. 13 magnetic and kinetic features are finally selected for the training of the machine learning model. Rank of each feature's Fisher score indicates that the duration of the south-directed interplanetary magnetic field that is larger than 10 nT and the increase

* 国家自然科学基金项目 (41731067, 41531073) 和中国科学院 "十三五" 信息化建设专项 (XXH13505-04) 共同资助

2018-06-05 收到原稿, 2019-01-31 收到修定稿

E-mail: ydye@spaceweather.ac.cn

of solar wind speed at the upstream shock or wave disturbance is closely related to the geoeffectiveness of ICME events, which is consistent with those former statistical results. The trained Radial Basis Function Support Vector Machine (RBF-SVM) can determine whether an ICME event could trigger moderate or stronger geomagnetic storms ($Dst \leqslant -50$ nT) effectively with an accuracy of 0.78±0.08. The results show that RBF-SVM can be used as a powerful tool in further analysis, and the better prediction of the geoeffectiveness of ICME will be obtained.

Key words Support Vector Machine, Interplanetary Coronal Mass Ejections, Geo-effectiveness

0 引言

行星际日冕物质抛射 (Interplanetary Coronal Mass Ejection, ICME) 是日冕物质抛射 (CME) 在行星际空间的对应部分. 通常 ICME 识别特征包括低的质子温度、低等离子体 β 值、双向电子流的出现以及铁和氧的异常电离态等 [1]. ICME 到达地球后, 其中的南向磁场分量在地球磁层向日侧诱发重联, 打开的磁力线使来自太阳风的动量、能量以及高能粒子注入磁层中, 带来剧烈的地磁扰动并引发地磁暴 [2]. 地磁暴的强度可以用 Dst 指数进行评估. Gonzalez 等 [3] 根据地磁暴期间 Dst 指数的不同, 对地磁暴的强度进行了分类. $Dst \leqslant -200$ nT 的磁暴为超强磁暴, $-200\,\mathrm{nT} < Dst \leqslant -100$ nT 为强磁暴, $-100\,\mathrm{nT} < Dst \leqslant -50$ nT 的磁暴为中等磁暴, $-50\,\mathrm{nT} / < Dst \leqslant -30$ nT 为弱磁暴.

在地磁活动预报中, 机器学习已被广泛应用. 这些应用主要以历史太阳风等离子体参数和磁场参数为主要输入, 结合历史 Dst 指数训练神经网络, 给出未来 1 h 或 4 h 不等的 Dst 指数预报 [4−6]. 这类研究关注点在于地磁暴中的地磁指数预报, 而ICME 地磁效应与哪些参数有关的研究工作主要还是基于统计分析 [7−9]. 这里引入机器学习思想和算法, 以 ICME 扰动期间的太阳风等离子体与磁场参数作为输入, 使用机器学习的特征选择算法评估各参量对 ICME 地磁效应的影响, 并训练支持向量机模型给出 ICME 扰动期间是否会有地磁暴的判断.

按照机器学习应用的完整流程, 本文选择用于机器学习模型输入的 ICME 参数和提取出的各个特征; 使用特征选择算法计算各特征的费舍尔分值 (Fisher Score) 并进行排序, 分析不同参量在引发地磁暴中的作用; 根据研究中使用的支持向量机相关概念以及训练的过程, 给出了支持向量机的测试结果, 讨论了选择不同 Dst 指数值作为正负样本划分所带来的各特征 Fisher Score 排序及训练得到的模型结果变化.

1 数据选取及特征提取

比较列表所提供的样本数量以及相关信息的丰富程度, 从已有的数个 ICME 列表中选择了 Richardson 和 Cane 的近地球 ICME 列表 [10,11] (下文中简称为 RC 列表) 作为 ICME 样本来源.

Richardson 和 Cane 在该列表中给出了利用 ACE 及 WIND 等多颗卫星根据探测到的太阳风等离子体参数以及磁场参数所证出的 ICME 事件以及每个事件的相关信息. RC 列表仍在不断更新并加入新认证的 ICME 事件. 这里选用列表中 1996 年至 2016 年的 483 个 ICME 样本及相关数据. RC 列表所提供的各个参量见表 1.

考虑到 BDE 数据和 BIF 数据缺失的比例接近 50%, 如果选择这些数据作为输入特征会带来 ICME 样本集过小的问题, 因此对这两项予以舍弃. 为了增强结果的可信度, 删除了 RC 列表中 “Qual.” 一栏中标注为 Weak 的事件. 另外, RC 列表中提供的磁场和等离子体信息有限. 为了更好地反映 ICME 扰动期间的太阳风等离子体和磁场状态, 参考 RC 列表中 (2) 参量提供的 ICME 事件的时间信息, 在 OMNIWeb[12] 中查询这期间的 1 h 时间分辨率的太阳风等离子体和磁场数据, 并计算该时间段内各参量的小时平均值. 选定的输入特征列于表 2. 表 2 中各特征的平均值为该特征在 ICME 事件期间的小时平均值. 最终形成含有 13 个特征和 433 个 ICME 事件的 ICME 样本集.

在选定的各特征中, 最大的太阳风速度可以达到

表 1 Richardson 和 Cane 的 ICME 列表所提供参量及含义

Table 1 ICME properties in Richardson and Cane's ICME list

序号	参量名	含义
1	Disturbance Y/M/D (UT)	地磁暴急始开始的时间
2	ICME Plasma/Field Start, (UT) End Y/M/D	基于等离子体与磁场数据估算的 ICME 事件开始/结束时间
3	Comp. Start, End (Hrs wrt. Plasma/Field)	异常太阳风结构/电离态持续时间
4	MC Start, End (Hrs wrt. Plasma/Field)	相对于 ICME 事件前缘/后缘的磁云事件的时间
5	BDE?	是否有双向超热电子流
6	BIF?	是否有双向高能离子流
7	Qual.	对 ICME 事件边界认证的可靠度
8	dV (km·s^{-1})	激波等波活动带来的 ICME 扰动开始时太阳风速度的增量
9	$V_{\rm ICME}$ (km·s^{-1})	平均 ICME 速度 (基于 (2) (3) 得到)
10	$V_{\rm max}$ (km·s^{-1})	ICME 事件中的太阳风速度最大值 (基于 (1) (3) 得到)
11	B (nT)	平均磁场强度 (基于 (2) (3) 得到)
12	MC?	是否为磁云事件
13	Dst (nT)	ICME 事件期间的 Dst 指数极小值
14	$V_{\rm transit}$ (km·s^{-1})	平均 1 AU 渡越速度 (基于 (15) 得到)
15	LASCO CME Y/M/D (UT)	ICME 事件所对应的最可能的 LASCO CME

表 2 选定的输入特征

Table 2 Selected input features

特征名称	含义
Mean magnetic field strength	平均磁场强度
Mean B_x	GSE 坐标下 x 方向平均磁场强度 B_x
Mean B_y	GSE 坐标下 y 方向平均磁场强度 B_y
Mean B_z	GSE 坐标下 z 方向平均磁场强度 B_z
Duration of $B_z < -10$ nT	ICME 事件中行星际磁场南北向分量 < -10 nT 的时间
Mean plasma β	平均等离子体 β 值
Mean α/proton ratio	平均 He/H 比值
Mean proton density	平均质子密度
Mean proton temperature	平均质子温度
Mean flow pressure	平均动压
Mean solar wind speed	平均太阳风速度
Maximum solar wind speed	ICME 事件期间的最大太阳风速度
Upstream increase in solar wind speed	ICME 扰动开始时太阳风速度的增量

1000 km·s^{-1} 或更高, 平均等离子体 β 值则大多分布于 0 至 1 之间. 为了消除不同特征数值极大差异带来的影响, 使用 Z 分值标准化方法 (Z-score Normalization) 处理 ICME 样本集内每个特征对应的数据. Z 分值体现的是以标准差为衡量标准所反映出来的特征值相对于均值的偏离程度. 标准差计算公式为

$$x_i = \frac{x'_i - \bar{x}'_i}{\sigma_i}. \tag{1}$$

其中, x'_i 为样本中特征 i 的值, $\bar{x}'_i$ 为全部样本中特征 i 的平均值, σ_i 为全部样本中特征 i 的标准差, x_i 为标准化处理后特征 i 的值. 经过处理后 x_i 的均值为 0, 标准差为 1, 所有数值均落在 $[-1, +1]$ 之间.

2 特征选择

完成数据标准化处理后, 以 ICME 事件期间的 Dst 指数极小值是否满足 $Dst \leqslant -50\,\mathrm{nT}$ 为标准, 对 ICME 样本集进行划分. 满足此条件的 ICME 事件样本选为正样本, 余下为负样本. 划分后得到 221 个正样本和 212 个负样本. 正样本数目与负样本数目比例接近于 1∶1, 比较均衡.

为了评估各特征在 ICME 事件引发中等及以上强度的地磁暴中所起作用的大小, 这里引入特征选择算法, 通过计算各特征的 Fisher Score[13] 对已有特征进行排序. 对于二分类问题 (结果为正类和负类), Fisher Score 由下式给出:

$$F(i) = \frac{n^+(\bar{x}_i^+ - \bar{x}_i)^2 + n^-(\bar{x}_i^- - \bar{x}_i)^2}{\dfrac{1}{(N-2)}\Big[\displaystyle\sum_{j=1}^{n_+}(x_{j,i}^+ - \bar{x}_i^+)^2 + \sum_{j=1}^{n_-}(x_{j,i}^- - \bar{x}_i^-)^2\Big]}. \tag{2}$$

其中, $\bar{x}_i^+$ 为正样本中特征 i 的平均值, $\bar{x}_i^-$ 为负样本中特征 i 的平均值, $\bar{x}_i$ 为全部样本中特征 i 的平均值, n^+ 为正样本的数目, n^- 为负样本的数目, N 为样本总数. Fisher Score 衡量的是一个特定特征在正样本和负样本中取值的差异程度. 如果某一特征的 Fisher Score 数值较大, 说明该特征在正负两组样本中的分布差异较大, 可用于有效区分正负样本. 研究发现, 具有高 Fisher Score 的特征能够显著影响 ICME 事件的地磁效应.

采用 Python 机器学习函数库 scikit-learn[14] 中的 Feature Scoring 函数组, 计算得到各特征 Fisher Score 的排序, 如图 1 所示.

Gonzalez 等 [3] 认为, 太阳风磁场 B_z 分量的持续南向更易引发地磁暴. Chi 等 [15] 的研究表明, 磁云事件的南向磁场分量更强, 也有更强的地磁效应. 研究证明含有激波的 ICME 事件的地磁效应较强 [16,17]. 由图 1 可以看出: 与磁云相关的物理参量如等离子体 β 值及 GSE 坐标系下 z 方向的磁场强度 Fisher Score 很高; ICME 事件中 $B_z < -10\,\mathrm{nT}$ 的持续时间的 Fisher Score 位列第二; 与 ICME 中激波等波活动强度有关的几个物理参量如 ICME 扰动开始时太阳风速度的增量、等离子体压强、GSE 坐标系下 x 方向磁场强度的 Fisher Score 较高. 同时, 由 Dst 指数的经验公式 [18] 可知, Dst 指数的变化与太阳风电场 E_y 直接相关. 太阳风电场 $E_y = vB_z$, 这里 v 为太阳风速度. 这解释了太阳风速度以及 B_z 得分较高的原因.

已提取的特征共有 13 个, 相对于 433 个的总样本数, 特征的数目较少. 因此, 以这些特征为基础训练得到的机器学习模型不易出现过拟合的现象, 在筛选特征集的过程中可以考虑剔除低分特征或者予以全部保留. 这里暂时保留所有特征进行模型训练.

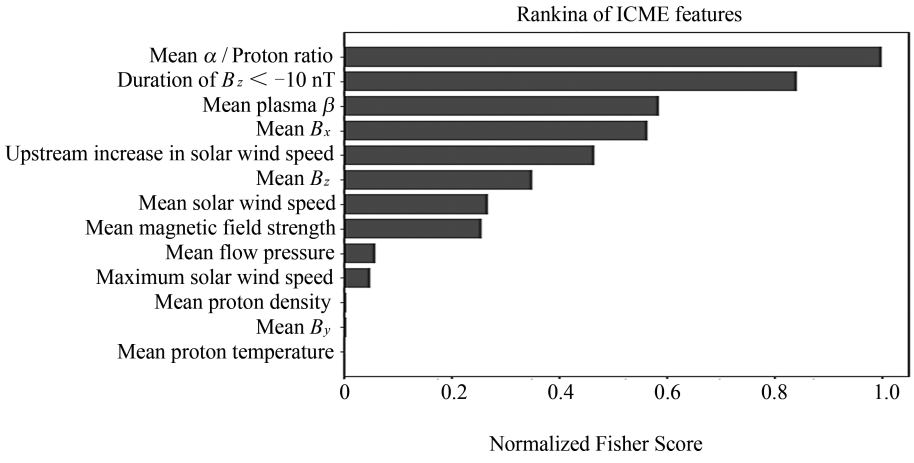

图 1 13 个特征的归一化 Fisher Score

Fig. 1 Normalized Fisher Score of all 13 features

3 支持向量机

选择 Vapnik 等 [19] 提出的支持向量机 (Support Vector Machine, SVM) 作为机器学习模型. 支持向量机求解分类问题的过程是在特征空间中寻找最优超平面的过程, 并且以软间隔 (Soft Margin) 最大化为最优超平面的划分目标. 软间隔支持向量机 (Soft Margin Support Vector Machine) 的训练过程等同于求解如下最优化问题:

$$\min \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i;$$

$$\text{s.t.} \quad y_i(wx_i + b) \leqslant 1 - \xi_i, \quad \xi_i \geqslant 0. \tag{3}$$

其中: ξ_i 为松弛变量, 用以平衡误分类点和最优超平面的距离; C 为惩罚因子, 用于平衡分类机对误分类的容忍程度. C 的数值越大意味着容忍程度越低, 相应会带来过拟合的风险. 软间隔线性支持向量机几何意义如图 2 所示.

对于非线性问题, 先使用核函数对特征空间进行变换, 将其投影至更高维度的空间后, 再使用线性支持向量机进行分类处理. 这里选择使用径向基函数 (Radial Basis Function, RBF) 作为的 SVM 进行训练. RBF 是一种类高斯核函数, 其表达式为:

$$K(x_i - x_j) = \exp(-\gamma\|x_i - x_j\|^2). \tag{4}$$

其中, $\|x_i - x_j\|$ 为两个样本点之间的欧氏距离. $\gamma > 0$, γ 值越大, 投影后的支持向量越少, 这会影响到训练

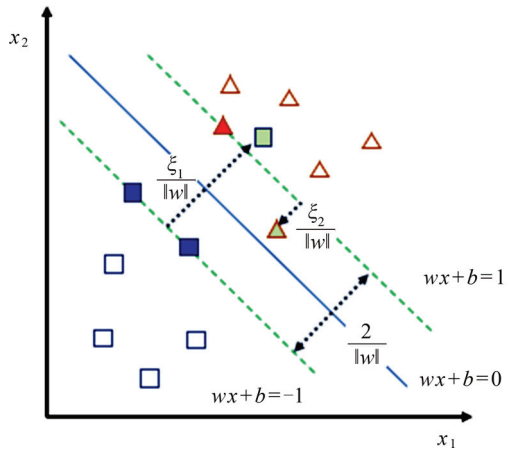

图 2 软间隔线性支持向量机几何意义

Fig. 2 Soft Margin Support Vector Machine

的速度和最终对训练数据的拟合情况. Amari 等 [20] 的研究发现, 测试集的最佳数目为 $S/\sqrt{2F}$, 其中 S 为总样本数, F 为总特征数. 按照本问题中 $S = 433$, $F = 13$ 计算可得, 训练集和测试集的比例为 80% 和 20%. 由于总样本数目以及正负样本数目均在数百个水平, 直接对 ICME 样本集进行 80%/20% 的划分无法有效保证训练集/测试集样本的随机性, 会影响到模型的泛化能力. 为了避免这个问题, 这里采用 k 分层交叉验证 (Stratified k-folds Cross-validation)[21] 划分训练集与测试集. k 分层交叉验证的方法随机将全部样本分为 k 等份, 并且每份均保持与 ICME 样本集中一致的正样本和负样本比例. 随机选择其中的 $k-1$ 份作为训练集, 余下的作为测试集. 这里 k 值的选择可以是从 2 至样本总数之间的任意值. Kohavi[22] 的研究表明, 在 k 取 10 的情况下, 可以有效减少模型的偏差和方差. 因此, 这里选择 10 层交叉验证方法来划分训练集和测试集.

为了确定 RBF-SVM 的两个关键参数 C 和 γ, 对 C 在 (0, 500] 区间内以 10 为步进取值, γ 在 (0, 0.2] 区间内以 0.005 为步进取值, 在此条件下寻找使 RBF-SVM 的真实技巧统计量 (True Skill Statistics, TSS)[23] 最高的一组 C 和 γ. 同时, 在训练过程中移除部分 Fisher Score 得分较低的特征. 在尝试了多种特征搭配之后发现, 移除平均质子温度后余下的 12 个特征可以取得最好的模型效果, 此条件下使 RBF-SVM 表现最好的参数为 $C = 300$, $\gamma = 0.035$.

4 结果与讨论

利用筛选出的特征集和择优选出的 C 和 γ 值, 对 RBF-SVM 进行训练并在测试集中测试. 通过比较 RBF-SVM 给出的判断结果与实际结果, 可以将测试集中的事件分为真正性 (True Positive, TP)、假正性 (False Positive, FP)、真负性 (True Negative, TN) 和假负性 (False Negative, FN) 四大类 (见表 3). 由此可以进一步计算得到准确率、精确率、召回率、F1 值和 TSS 值.

这些评价标准的计算公式如下.

准确率 (Accuracy)

$$A = \frac{N_{\text{TP}} + N_{\text{TN}}}{N_{\text{total}}}. \tag{5}$$

精确率 (Precision)

$$P = \frac{N_{\rm TP}}{N_{\rm TP} + N_{\rm FP}}. \tag{6}$$

召回率 (Recall)

$$R = \frac{N_{\rm TP}}{N_{\rm TP} + N_{\rm FN}}. \tag{7}$$

精确率和召回率的调和平均值

$$F_1 = \frac{2N_{\rm TP}}{2N_{\rm TP} + N_{\rm FP} + N_{\rm FN}}. \tag{8}$$

其中, $N_{\rm TP}$, $N_{\rm FP}$, $N_{\rm TN}$, $N_{\rm FN}$ 和 $N_{\rm Total}$ 分别为 TP, FP, TN, FN 类事件的数量以及各类别事件的总数.

准确率给出了 RBF-SVM 做出正确判断的比例. 精确率衡量的是 RBF-SVM 判断为存在中等强度及以上地磁暴的 ICME 事件中有多少事件确实引发相应强度的地磁暴, 召回率给出中等强度及以上地磁暴的 ICME 事件有多少被正确地预报.

TSS 可用于衡量 RBF-SVM 能否正确做出判断. TSS 取值范围为 −1∼1, 越接近于 1 表明模型的效果越好. TSS 的计算公式为

$$S_{\rm TSS} = \frac{N_{\rm TP}N_{\rm TN} - N_{\rm FP}N_{\rm FN}}{(N_{\rm TP} + N_{\rm FN})(N_{\rm FP} + N_{\rm TN})}. \tag{9}$$

最终模型的结果列于表 4.

各评价参数的结果表明, 将剔除平均质子温度之后剩余的 12 个特征作为输入特征, 使用 RBF-SVM 可以有效给出 ICME 扰动是否会带来中等及以上强度地磁暴的判断.

更进一步可以将区分正负样本的标准修改为 Dst 指数是否 $\leqslant -100\,{\rm nT}$, 即 ICME 是否会带来强地磁暴及超强地磁暴. 在这种划分标准下, 可将 ICME 样本集划分为 50 个正样本与 269 个负样本. 重新计算 Fisher Score 并对所有特征进行排序, 进一步检验训练得到的分类机的表现, 得到的 Fisher Score 排序如图 3 所示.

表 3 观测结果与预报结果

Table 3 Observation and forecast results

		观测结果	
		是	否
预报结果	是	真正性 (TP)	假正性 (FP)
	否	假负性 (FN)	真负性 (TN)

表 5 中排序序号越小表明 Fisher Score 得分值越高. 对比表 5 的排序可以发现, 在以 Dst 指数为 $-100\,{\rm nT}$ 作为训练集正负样本的划分判据时, 引发强磁暴或超强磁暴的 ICME 在最大速度、GSE 坐标系下 x 方向的磁场强度、平均磁场强度等方面得分排序变化很大, 但是平均质子数密度、GSE 坐标系下 y 方向的磁场强度和南向磁场等特征得分排序的变化不大. 在以 Dst 指数为 $-100\,{\rm nT}$ 作为样本集划分判据时, 与磁云有关的特征并不能很好地区分正负

表 4 RBF-SVM 结果

Table 4 Results of RBF-SVM model

准确率	0.78±0.08
精确率	0.81±0.10
召回率	0.76±0.11
F1 值	0.78±0.10
TSS 值	0.56±0.15

表 5 以 Dst 指数为 $-50\,{\rm nT}$ 和 $-100\,{\rm nT}$ 为标准划分正负样本集计算 Fisher Score

Table 5 All 13 features' Fisher Score as the index of Dst is equal to $-50\,{\rm nT}$ and $-100\,{\rm nT}$

特征名	排序	
	$-50\,{\rm nT}$	$-100\,{\rm nT}$
Mean α/proton ratio	1	12
Duration of $B_z < -10\,{\rm nT}$	2	8
Mean plasma β	3	5
Mean B_x	4	2
Upstream increase in solar wind speed	5	4
Mean B_z	6	9
Mean solar wind speed	7	10
Mean magnetic field strength	8	3
Mean flow pressure	9	7
Maximum solar wind speed	10	1
Mean proton density	11	11
Mean B_y	12	13
Mean proton temperature	13	6

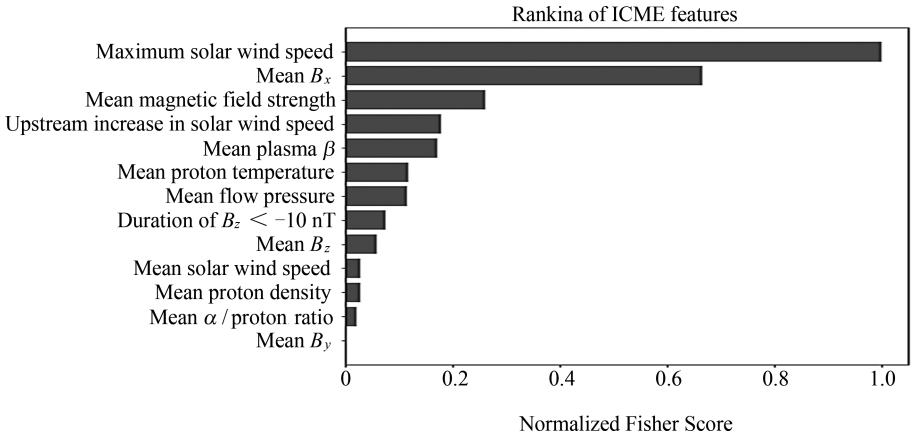

图 3 以 −100 nT 作为划分标准得到 13 个特征的 Fisher Score

Fig. 3 Normalized Fisher Score of all 13 features as the classification criteria is −100 nT

样本. 这说明磁云事件在引发强磁暴及超强磁暴过程中的作用并不显著.

针对强地磁暴和超强地磁暴的研究表明, 多个 ICME 会在行星际空间发生相互作用. 在这种相互作用过程中, 后一个高速 ICME 的激波会压缩前一个 ICME 的磁场区域并带来加速 [24]. 与含有孤立 ICME 的事件以及包含多个 ICME 但不存在相互作用的事件相比, 这种存在相互作用的事件有着更强的地磁效应 [25]. ICME 的最大速度、GSE 坐标系下 x 方向的磁场强度以及 ICME 扰动结构开始时太阳风的速度增量等特征能够在一定程度上反映出这种相互作用, 这也是这些特征得分较高的原因. 但是, 本文选用的特征集合中仍然缺少能够更加准确描述一次 ICME 事件中是否包含多个 ICME 以及多个 ICME 之间是否存在相互作用的特征. 此外, 以 −100 nT 为判据划分得到的正负样本不均衡性较高 (正样本/负样本比例接近 1/52). 在这种划分下, RBF-SVM 给出 ICME 事件是否会引发强磁暴以及超强磁暴的判断效果稍差, TSS 数值为 0.38±0.15.

另外, 从表 5 的排序变化可以发现, 在以 −50 nT 作为训练集正负样本划分判据时, 平均 He/H 比值的 Fisher Score 得分排序为 1. 但是, 以 −100 nT 作为划分判据时, 其排序为 12, 变化很大. 由于各特征的 Fisher Score 数值和样本集中 ICME 的特性直接相关, 因此该变化背后的物理含义有待于通过进一步分析各特征在对地效应中所起的作用来进行深入研究. 在未来的工作中, 需要整理含有更多不同类型 ICME 事件的样本集, 并且引入能够描述 ICME 或对多个 ICME 事件进行细致划分的特征量, 寻找更加完备的输入特征集合.

参考文献

[1] ZURBUCHEN T H, RICHARDSON I G. In-situ solar wind and magnetic field signatures of interplanetary coronal mass ejections [J]. *Space Sci. Rev.*, 2006, **123**(1/2/3): 31-43

[2] LAKHINA G S, TSURUTANI B T. Geomagnetic storms: historical perspective to modern view [J]. *Geosci. Lett.*, 2016, **3**(1): 5

[3] GONZALEZ W D, JOSELYN J A, KAMIDE Y, *et al.* What is a geomagnetic storm [J]. *J. Geophys. Res.*, 1994, **99**: 5771-5792

[4] WU J G, LUNDSTEDT H. Geomagnetic storm predictions from solar wind data with the use of dynamic neural networks [J]. *J. Geophys. Res.*: *Space Phys.*, 1997, **102**(A7): 14255-14268

[5] BALA R, REIFF P. Improvements in short-term forecasting of geomagnetic activity [J]. *Space Weather*, 2012, **10**(6). DOI:10.1029/2012SW000779.

[6] WATANABE S, SAGAWA E, OHTAKA K, *et al.* Prediction of the *Dst* index from solar wind parameters by a neural network method [J]. *Earth Planets Space*, 2014, **54**(12): 1263-1275

[7] WANG Y M, YE P Z, WANG S, *et al.* A statistical study on the geoeffectiveness of Earth-directed coronal mass ejections from March 1997 to December 2000 [J]. *J. Geophys. Res.*, 2002, **107**(A11). DOI:10.1029/2002JA009244.

[8] GOPALSWAMY N, YASHIRO S, XIE H, *et al.* Properties and geoeffectiveness of magnetic clouds during solar cycles 23 and 24 [J]. *J. Geophys. Res.*: *Space Phys.*, 2015, **120**(11): 9221-9245

[9] LAWRANCE M B, SHANMUGARAJU A, MOON Y J, *et al*. Relationships between interplanetary coronal mass ejection characteristics and geoeffectiveness in the rising phase of solar cycles 23 and 24 [J]. *Sol. Phys.*, 2016, **291**(5): 1547-1560

[10] CANE H V, RICHARDSON I G. Interplanetary coronal mass ejections in the near-Earth solar wind during 1996-2002 [J]. *J. Geophys. Res.*, 2003, **108**(A4). DOI:10.1029/2002JA009817.

[11] RICHARDSON I G, CANE H V. Near-earth interplanetary coronal mass ejections during solar cycle 23 (1996–2009): catalog and summary of properties [J]. *Sol. Phys.*, 2010, **264**(1): 189-237

[12] MATHEWS G J, TOWHEED S S. NSSDC OMNIWeb: The first space physics WWW-based data browsing and retrieval system [J]. *Comp. Networks ISDN Syst.*, 1995, **27**(6): 801-808

[13] GU Q Q, LI Z H, HAN J W. Generalized Fisher Score for Feature Selection [C]//Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence, 2011: 266-273

[14] PEDREGOSA F, VAROQUAUX G, GRAMFORT A, *et al*. Scikit-learn: machine learning in Python [J]. *J. Mach. Learn. Res.*, 2011, **12**: 2825-2530

[15] CHI Yutian, SHEN Chenglong, WANG Yuming, *et al*. Statistical study of the interplanetary coronal mass ejections from 1995 to 2015 [J]. *Sol. Phys.*, 2016, **291**(8): 2419-2439

[16] WANG Yuming, YE Pinzhong, WANG Shui. An interplanetary origin of great geomagnetic storms: multiple magnetic clouds [J]. *Chin. J. Geophys.*, 2004, **47**(3): 417-423

[17] TSURUTANI B T, GONZALEZ W D, TANG F, *et al*. Great magnetic storms [J]. *Geophys. Res. Lett.*, 1992, **19**(1): 73-76

[18] BURTON R K, MCPHERRON R L, RUSSELL C T. An empirical relationship between interplanetary conditions and Dst [J]. *J. Geophys. Res.*, 1975, **80**(31): 4204-4214

[19] VAPNIK V, GOLOWICH S E, SMOLA A J. Support vector method for function approximation, regression estimation and signal processing [C]//Proceedings of the Advances in Neural Information Processing Systems, 1997: 281-287

[20] AMARI S, MURATA N, MULLER K R, *et al*. Asymptotic statistical theory of overtraining and cross-validation [J]. *IEEE Trans. Neural Netw.*, 1997, **8**(5): 985-996

[21] REFAEILZADEH P, TANG L, LIU H. Cross-validation [R]. Encyclopedia of Database System, US: Springer. DOI: 10.1007/978-0-387-39940-9_565

[22] KOHAVI R. A study of cross-validation and bootstrap for accuracy estimation and model selection [C]//Proceedings of the 14th International Joint Conference on Artificial Intelligence – Volume 2. Quebec: Morgan Kaufmann Publishers Inc, 1995: 1137-1143

[23] FLUECK J. A study of some measures of forecast verification [C]//Proceedings of the Preprints, 10th Conference on Probability and Statistics in Atmospheric Sciences. Edmonton: Alberta, Am. Meteor. Soc., 1987: 69-73

[24] SHEN Fang, WANG Yuming, SHEN Chenglong, *et al*. On the collision nature of two coronal mass ejections: a review [J]. *Sol. Phys.*, 2017, **292**(8): 104

[25] SHEN Chenglong, CHI Yutian, WANG Yuming, *et al*. Statistical comparison of the ICME's geoeffectiveness of different types and different solar phases from 1995 to 2014 [J]. *J. Geophys. Res.*: *Space Phys.*, 2017, **122**(6): 5931-5948