

基于显著性特征的大冰雹识别模型*

王萍 潘跃†

(天津大学电气与自动化工程学院, 天津 300072)

(2012年8月31日收到; 2012年11月13日收到修改稿)

为了解决目前气象业务中由数字化雷达系统提供的强冰雹指数对强冰雹空报率过高的问题, 依据概念模型对强冰雹回波单体特点的描述, 设计并实现了“悬垂度”等多个特征提取算法, 为了校验这些特征以及传统的强冰雹指数是否能从各自的角度突显强冰雹单体自身特质, 将较易与其混淆的短时强降水单体作反例进行统计校验, 结果表明它们在两类样本之间的差异是显著的. 以此为基础, 选用基于径向基核函数的非线性支持向量机得到强冰雹识别模型, 并在一种尺度变换的基础上, 将待测样本远离最优分类超平面的相对程度定义为新冰雹指数. 实验表明, 本文方法较目前业务上普遍使用的冰雹指数法具有更高的击中率, 同时空报率大大降低.

关键词: 强冰雹识别, 风暴单体的悬垂特征, 显著性差异特征, 支持向量机

PACS: 92.60.Nv, 92.60.Qx, 92.60.Wc

DOI: 10.7498/aps.62.069202

1 引言

冰雹灾害是由强对流天气系统引起的一种剧烈的气象灾害, 它出现的范围较小, 时间短促, 但来势猛烈, 并常常伴随着狂风、强降水、急剧降温等阵发性灾害性天气过程. 新一代天气雷达是监测甚至预测包括冰雹在内的强对流天气的重要平台, 其中最具代表性的是美国于 1996 年规模化应用于全美的配有较丰富应用软件的多普勒天气雷达气象业务系统, 该系统所提供的冰雹指数与所替代的冰雹指数相比, 使美国强冰雹预警的击中率维持在 70% 的同时, TS 评分从 26% 增至 42%^[1-3]. 目前, 新一代天气雷达及其应用软件算法与美国基本相同, 该系统每隔 6 min 提供一组多个探测仰角下的回波图像, 特别是其反射率因子图和径向速度图能够直观展现对流云团的结构、形态以及对流场的分布, 从而促进了人们对短时强降水、强冰雹、灾害性大风等不同类型的强对流天气特点及其演变规律的认识. 就强冰雹而言, 人们认识到:

1) 凡发生冰雹的云团, 至少有一部分越过了环境融化层, 即零度层, 越过的部分越多、越过的程度越高、零度层的高度 (与地表温度有关) 越低, 降

强冰雹的可能性越大, 反映在雷达回波的反射率因子图上, 高仰角的冰雹回波强度会持续很高;

2) 在强冰雹云团回波的反射率因子图上, 通常会呈现一种“悬垂”结构, 即位于中高层的较强回波区域会“探”到其低层较强回波区域之“外”, 换言之, 即在底层弱回波区域之上“悬垂”着一部分较强的回波体^[4];

3) 强冰雹云团的低层回波出现明显的梯度不均等结构, 即从回波高强度中心递减到低强度边界所经距离的差异极大, 导致强回波中心偏倚于回波图像的一侧^[5], 上述悬垂现象就多发生于该低层回波图像的极高或较高梯度的方位;

4) 与短时强降水回波相比, 强冰雹的高强度回波范围所占比例更大^[6].

另外, 还认识到一些评定强冰雹的充分条件, 即回波的反射率因子图像一旦出现“三体散射”或“钩”、径向速度图像一旦出现“强中气旋”, 则肯定出现强冰雹^[7].

目前, 气象上普遍使用的业务系统提供的强冰雹指数算法是考虑上述特点 1) 设计的, 因为实际上许多短时强降水回波也会有一部分甚至大部分扩展到零度层高度以上, 因此, 基于这一强冰雹指

* 天津市自然科学基金 (批准号: 09JCYBJC07500) 和公益性行业 (气象) 科研专项基金 (批准号: GYHY200706004) 资助的课题.

† 通讯作者. E-mail: panyuetj@gmail.com

数的强冰雹误报(空报)率会很高,据美国给出的统计,空报率高达 49%^[8]. 在国内,由于这一过高(经常会高于 49%)的空报率,许多气象台只能将其作为强冰雹的起报条件,有的甚至直接用其辅助预报“短时强降水”^[9]. 为此,人们展开了一些积极的研究和探索,其中,所提出的“指标加权法”^[10]增加了液态水含量、中气旋、回波形态、三体散射等多个预报因子,通过对各预报因子分段赋值再求和的方法,获得对冰雹的 TS 评分的测试结果达到 75%;“相似演变聚类法”^[11]虽然还是主要选择回波高度及强度类指标作因子,但通过与已发生的冰雹过程贴近度的计算,使对冰雹的 TS 评分达到 65% 等.

本文在考虑冰雹指数的同时,新增了反映上述特点 2)—4) 的特征提取算法,并以短时强降水回波为反例,在对诸项特征显著性分析的基础上,联合使用这些特征训练出强冰雹识别模型. 推出一种新型冰雹指数,最后对新的冰雹指数与传统的冰雹指数进行了对比测试.

2 强冰雹回波的特征提取

根据气象理论和大量观测经验可知,发生强对流天气时的多个仰角的多普勒雷达回波图中会出现显著的强“单体”,即在多个仰角的反射率图像中出现以大于等于 50 dBZ 的区域为核,从“核”向外反射率逐渐由高(50 dBZ)到低(25 dBZ)过渡的区域. 本文首先以多仰角的反射率回波单体为对象,根据上小节指出的冰雹回波特点,构建用于鉴别冰雹回波的系列特征.

2.1 多仰角单体回波的匹配

在强对流可能出现的时段,多普勒天气雷达多采用 9 仰角的扫描模式,通过 9 张回波强度图像捕获对流体的立体结构信息,因此,在多单体共存的通常情况下,需要确定各单体在最多 9 张仰角图中的匹配关系.

2.1.1 摒弃高度信息的重叠率规则

图 1 是雷达通过两仰角方式探测垂直线型对象 A 的示意图. 在探测仰角 α_1 下的锥面探测图像 I_1 中,探测 A 为点 $P_1(r_1, \theta_1, h_1)$,同理,在探测仰角 α_2 下,探测 A 为点 $P_2(r_2, \theta_2, h_2)$,将点 P_1 和点 P_2 向水平面投影,得到 $P'_1(r'_1, \theta'_1)$ 和 $P'_2(r'_2, \theta'_2)$,其中

$$\begin{cases} r'_1 = r_1 \cos \alpha_1 \\ r'_2 = r_2 \cos \alpha_2 \end{cases}, \quad (1)$$

容易推知: $\theta_1 = \theta_2, r'_1 = r'_2$,即点 P'_1 和点 P'_2 完全重合. 推而广之,若对象 A 为任意水平截面形态完全相同的垂直物体,则在其锥面图像 I_1 和 I_2 的探测结果虽不相同 ($s_1 \neq s_2$),但其水平投影依然完全重合 ($s'_1 = s'_2$). 一个强对流云体几乎是垂直的,但不同高度的形态、面积大小会有差异,强冰雹的“悬垂”结构还会使这种差异变大,因此,在对同一单体的多仰角探测图像进行匹配时,将重合条件放宽至小于 1 的一个比率 ρ ,即来自相邻两仰角锥面中对象的水平投影区域重叠率超过 ρ ,则认为两者属于同一单体.

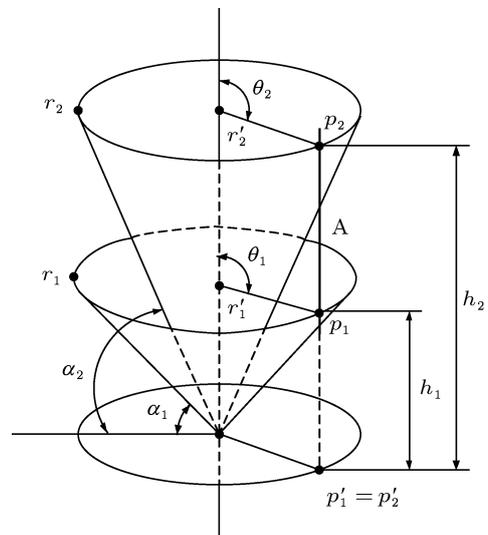


图 1 垂直对象水平投影示意图

2.1.2 单体的纵向匹配算法

1) 对投影后的九个仰角反射率因子图 I_1, I_2, \dots, I_9 进行二值化处理,保留 30 dBZ 以上区域并进行形态学运算,除去孤立块、填补孔洞、去除毛刺;

2) 取每个连通的区域最小外包矩形 M_{pi_p} , $p = 1, \dots, 9, i_p = 1, \dots, n_p$;

3) 将高层 M_{pi_p} 逐层逐个与低层 M_{qj_q} 进行比较,若对任意的 $q > 0, q < p < 10$ 都存在一个 M_{qj_q} 使得

$$M_{pi_p} \cap M_{qj_q} / M_{pi_p} > M_\alpha, \quad (2)$$

其中, M_α 为常数,这里取 0.2,记 $M_k = M'_k = \{M_{1j'_1}, M_{2j'_2}, \dots, M_{pj'_p}\}$, $k = 1, \dots, n$ (n 为单体数);并将满足 (2) 式的 M_{qj_q} 作标记;

4) 用未进行比较或未被标记的 M_{pi_p} 重复 3),直到所有 M_{qj_q} 都被标记;

5) 若最低仰角反射率因子图 I_1 中有未被标记

的 M_{1i_1} , 直接将 M_{1i_1} 记为 M_k ;

6) 若低层的同一连通区域的最小外包矩形被多次标记, 如图 2 所示, 低层连通区域的最小外包矩形 a 与其上层两个连通区域的最小外包矩形 b 和 c 都匹配, 此时 a 被标记两次, 则取消相关标记, 提高分割阈值, 对多次标记的矩形框内区域重新进行阈值分割, 返回 4), 直至各层各单体都被一一匹配。

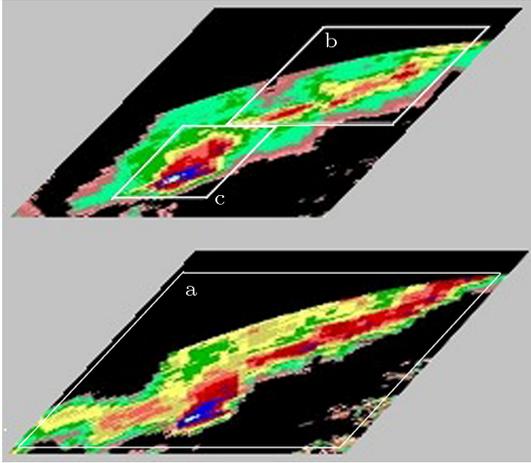


图 2 低层同一区域被上层的多区域标记示例

2.2 传统冰雹指数简介

首先, 根据单体在某一仰角反射率图像上的强度值, 计算与地面冰雹灾害潜势密切相关的“冰雹动能” $E^{[12]}$

$$E = 5 \times 10^{-6} \times 10^{0.084Z} W(Z), \quad (3)$$

其中, Z 为单体在该仰角图像上的最大反射率值, $W(Z)$ 为权重

$$W(Z) = \begin{cases} 0 & Z \leq Z_L \\ \frac{Z - Z_L}{Z_U - Z_L} & Z_L < Z < Z_U \\ 1 & Z \geq Z_U \end{cases}, \quad (4)$$

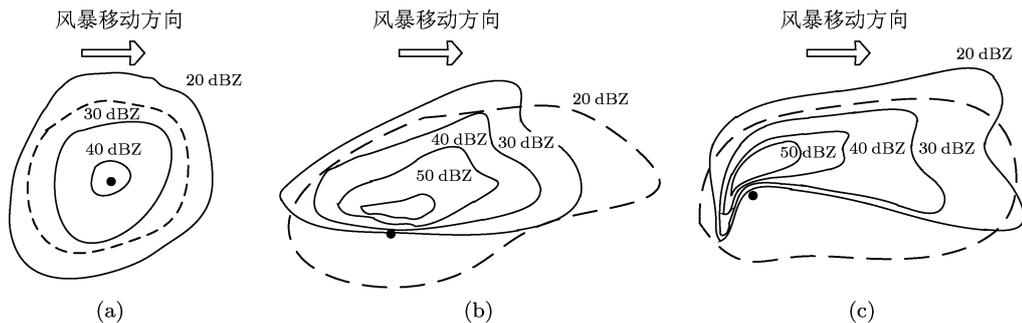


图 3 风暴单体反射率因子垂直结构示意图 (a) 单体非强风暴; (b) 单体强风暴; (c) 超级单体风暴

该权重使最大反射率强度与冰雹动能正相关, 并使最大反射率值低于 Z_L 的回波区域不对“冰雹动能”做出贡献。

再求环境融化层 (零度层) H_0 以上的冰雹动能加权通量:

$$SHI = 0.1 \int_{H_0}^{H_T} W_T E dH, \quad (5)$$

其中权重函数

$$W_T(H) = \begin{cases} \frac{H - H_0}{H_{-20} - H_0} & H_0 < H < H_{-20} \\ 1 & H \geq H_{-20} \end{cases} \quad (6)$$

表明, 单体跃出零度层高度的程度 $(H - H_0)$ 与总冰雹动能正相关, 同时认为, 低于 H_0 的冰雹动能不会对降雹做出贡献。该冰雹动能加权通量就是传统的强冰雹指数, 在用其进行强冰雹监测和预警时, 通常使用根据美国 Oklahoma 州和 Florida 州风暴实况的相关分析得出的预警阈值函数:

$$WT = 57.5H_0 - 121, \quad (7)$$

即若 $SHI > WT$, 则进行强冰雹预警^[13]。

可见, 传统强冰雹指数属于反映强冰雹回波特点 (1) 的一种特征。

2.3 单体的悬垂度特征

根据气象预报经验, 当对流风暴的低仰角反射率因子图具有宽阔的弱回波区或有界弱回波区, 且在它们上方存在强反射率因子核时, 最有利于强冰雹的发生。图 3(a)—(c) 给出了单体非强风暴、单体强风暴和超级单体风暴反射率因子垂直结构示意图, 实线为单体低层回波强度的等值线, 虚线为中层回波强度大于 20 dBZ 的轮廓线, 实心点为高层最大回波强度处。可见, 具有冰雹潜势的单体强风暴 (图 3(b)) 和超级单体风暴 (图 3(c)), 其反射率因子结构在低层高梯度侧呈现强回波“悬垂”。

2.3.1 区域平均梯度

为定位出悬垂关注区, 需要寻找强风暴在低层回波图像中反射率快速及较快速递减的区域, 也就是高梯度区.

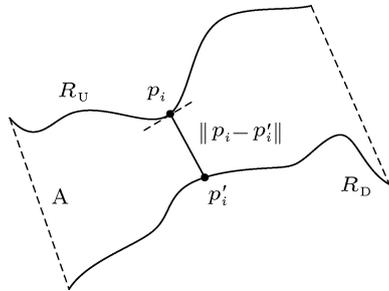


图4 从 R_U 过渡到 R_D 的局部区域示意图

图4是某单体低层回波从高反射率 R_U 等值线过渡到低反射率 R_D 等值线的局部情况, 在关注区域 A 中, 从点 p_i 到 p'_i 的梯度

$$G(p_i, p'_i) = \frac{R_U - R_D}{\|p_i - p'_i\|}, \quad (8)$$

因为 $(R_U - R_D)$ 为常数, $G(p_i, p'_i)$ 与 $\|p_i - p'_i\|$ 成反比, 为降低运算时间开销, 特别定义区域 A 从 R_U 过渡到 R_D 的平均梯度为

$$\bar{T}_A = \frac{1}{\frac{1}{N_A} \left(\sum_{i=1}^{N_A} \frac{1}{G_i} \right)}, \quad (9)$$

即

$$\bar{T}_A = \frac{1}{\frac{1}{N_A} \left(\sum_{i=1}^{N_A} \frac{\|p_i - p'_i\|}{R_U - R_D} \right)}$$

$$= \frac{R_U - R_D}{\frac{1}{N_A} \sum_{i=1}^{N_A} \|p_i - p'_i\|} \approx \frac{R_U - R_D}{\frac{1}{N_A} S_A}, \quad (10)$$

其中, S_A 为区域 A 的面积. 可见区域 A 的平均梯度与其平均宽度成反比, 若设区域 A 的长度为常数, 则区域 A 的平均梯度与其面积成反比, 因此, 寻找最大梯度区, 等价于寻求平均宽度最小区域或最小面积区域.

2.3.2 悬垂区域关注系数

考虑到出现“悬垂”的高梯度区具有一定的长度, 且高梯度区一般不会突变到最低或次低梯度区, 首先对单体进行处理, 保留 R_U 至 R_D 的区域 (缺省值: $R_U = 45 \text{ dBZ}$, $R_D = 30 \text{ dBZ}$), 如图 5(a)—(c) 所示, 然后以单体的强反射率核心为中心作射线将单体分为 8 份 A_1, A_2, \dots, A_8 (见图 5(d)), 分别求出它们的面积 S_1, S_2, \dots, S_8 . 一般情况下, 区域面积越小, 其平均梯度越大, 将其按照从小到大的顺序排序, 考虑到较大梯度区域不一定刚好落入一个分区内以及悬垂一定不会出现最低及次低梯度区之上, 设悬垂区域关注系数 $\beta = \{0.3, 0.3, 0.2, 0.1, 0.05, 0.05, 0, 0\}$.

2.3.3 悬垂度提取算法

1) 一个匹配好的风暴云团单体 k , 经多仰角匹配后得到外包矩形集合 $M_k = \{M_{k1}, M_{k2}, \dots, M_{kN}\}$, 其中 $k = 1, \dots, K$, K 为一个时刻回波图像中出现的单体数, 对每一个 k 执行以下过程 (为简单起见, 略去 k 不记).

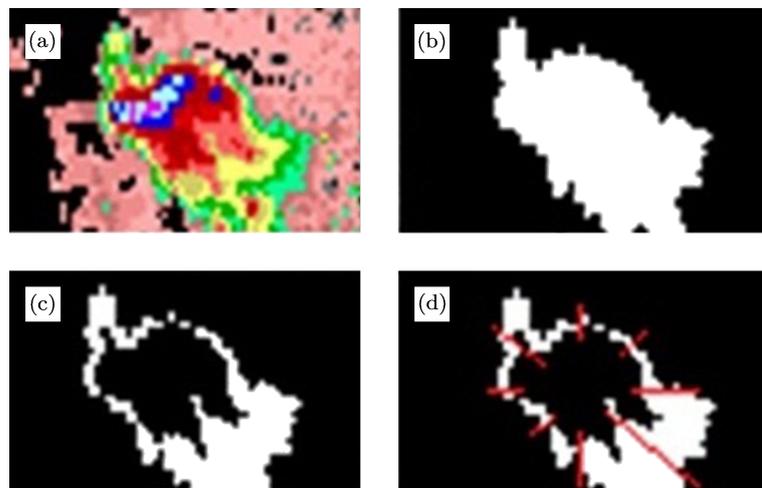


图5 八射线区域分割示意图 (a) 单体低仰角反射率图; (b) 保留反射率值大于等于 R_D 的区域; (c) 减去 (b) 区域中反射率值大于等于 R_U 的区域; (d) 用八条射线将 (c) 区域等分

2) 对 M_1 , 求取梯度区

$$\Delta A = A|_{R \geq R_D} - A|_{R > R_U}, \quad (11)$$

将 ΔA 用射线等分成 8 份, 并求取相应的面积 ΔS_i , 对八等份区域根据 ΔS_i 的大小按升序编号 n , $n = 1, \dots, 8$.

3) 对 $M_j, j = 1, \dots, N$, 做与 2) 相同的八等份划分, 得到 M_{ji} 中的有效回波面积 S_{ji} , 并赋予编号, 即 $S_{ji}(n)$.

4) 计算第 m 层回波对 $m - 1$ 层回波的局部悬垂度

$$q_m = \sum_{i=1}^8 \beta_n \frac{S_{mi}(n) - S_{mi}(n) \cap S_{(m-1)i}(n)}{S_{mi}(n) \cup S_{(m-1)i}(n)}, \quad (12)$$

$$m = 2, \dots, N.$$

5) 计算总悬垂度

$$G = \sum_{m=1}^N q_m. \quad (13)$$

2.4 高回波比

雹暴通常与大片的强雷达回波相联系, 高回波比用于反映此项特点, 定义为 50 dBZ 及以上区域占单体 30 dBZ 及以上区域的比重. 该特征不涉及高度信息, 因此可借助单体匹配成功的外包矩形集合 $M = \{M_1, M_2, \dots, M_N\}, N \leq 9$ 求得.

设矩形区域 M_i 中, 取值大于等于 50 dBZ 的像素数为 N_{i-50} , 取值大于等于 30 dBZ 的像素数为 N_{i-30} , 则定义高回波比特征如下:

$$\rho_H = \frac{\sum_{i=1}^N N_{i-50}}{\sum_{i=1}^N N_{i-30}}. \quad (14)$$

2.5 峰度特征

图 6(a), (b) 是某个短时强降水单体和某个强冰雹单体的低仰角 (0.5°) 反射率因子图, 图 7 是它们的强度分布直方图. 直观地看, 两者有着高强度值成分略少、低强度比例相差不大、单侧尾部较厚的分布规律. 基于分布直方图的“峰度”是一个突出“厚重尾部”的 4 阶统计量, 定义式如下:

$$\mu_4 = \frac{\sum_{i=1}^N (Y_i - \bar{Y})^4}{(N - 1)s^4}, \quad (15)$$

其中, \bar{Y}, s 是其均值和标准差. 因为强冰雹的最高反射率普遍高于短时强降水, 或者说强冰雹的较高反射率所占份额普遍高于短时强降水, 因此强冰雹的

均值普遍高于短时强降水的均值, 从而导致强冰雹的“峰度”普遍高于短时强降水的“峰度”.

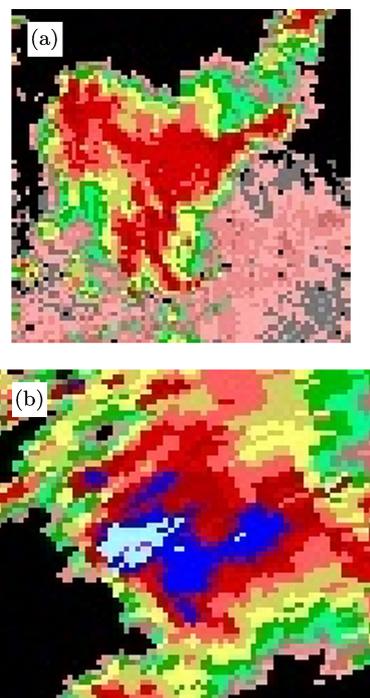


图 6 强冰雹与短时强降水反射率图对比 (a) 短时强降水; (b) 强冰雹

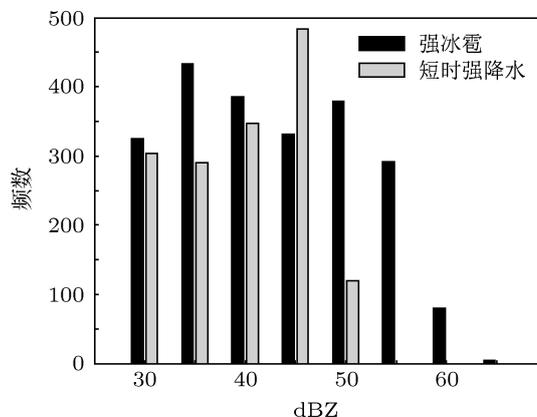


图 7 强冰雹单体和短时强降水单体反射率强度分布直方图

另外, 本文还设计了有效厚度特征 ΔH , 即大于 45 dBZ 的单体核的最大高度与零度层高度 H_0 之差:

$$\Delta H = \max\{H_{R > 45 \text{ dBZ}}\} - H_0. \quad (16)$$

2.6 特征的显著性分析

为检验所提特征的显著性, 随机选择来自 14 过程的强冰雹单体 60 个和来自 12 个过程的短时强降水单体 52 个进行测试, 形成基于单个特征的直方图分布, 如图 8 所示.

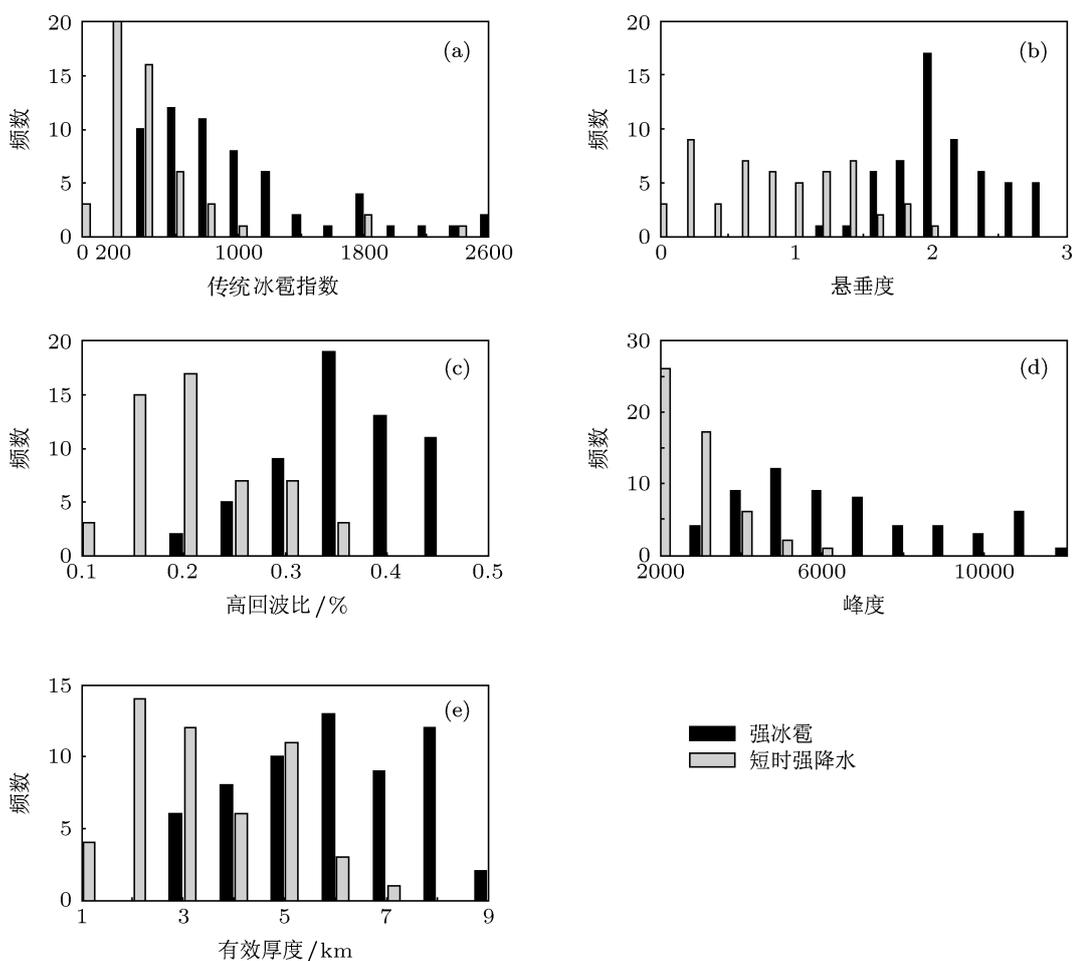


图 8 关于强冰雹样本和短时强降水样本的各特征统计直方图 (a) 传统冰雹指数统计直方图; (b) 悬垂度统计直方图; (c) 高回波比统计直方图; (d) 峰度统计直方图; (e) 有效厚度统计直方图

设用于测试特征有效性的强冰雹样本和短时强降水样本分别来自两个来自正态分布的总体, 且方差相同. 下面根据测试数据 (图 8) 推断由相关特征描述的这两个总体有无显著性差异, 并由此推断所提特征的有效性, 即假设强冰雹均值等于短时强降水均值, 使用 (17) 式的服从 t 分布的统计量展开置信水平 $(1 - \alpha)$ 的检验. 计算结果如表 1 所示.

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t(n_1 + n_2 - 2) \quad (17)$$

式中, \bar{x}, \bar{y} 分别为强冰雹和强降水样本的特征均值, S_1^2, S_2^2 分别为对应的方差.

取显著性水平 $\alpha = 0.01$, 则 $t_{\alpha/2}(n_1 + n_2 - 2) = t_{0.005}(112)$, 查表得 $t_{0.005}(112) < 2.660$. 由于 5 个特征计算结果均使 $t_i > 2.66 > t_{0.005}(112), i = 1, \dots, 5$, 因此推翻原假设, 认为本文中所选各项特征在短时

强降水和强冰雹之间均具有显著性差异.

表 1 5 个特征的单因子显著性差异检验

特征	传统强冰雹指数	悬垂度	高回波比	峰度	有效厚度
t	5.576	14.145	12.049	10.536	8.314

3 强冰雹识别模型

上小节所提 5 项特征均是站在强冰雹回波特点的角度上设计和实现的. 图 8 和表 1 所展现的实测结果表明, 它们在短时强降水和强冰雹两类风暴之间的统计性差异还是很明显的. 另外, 根据气象理论, 零度层高度也是左右冰雹强弱的重要特征. 因此, 本文特别将短时强降水这种极易与强冰雹混淆的单体作为反例, 训练基于这 5 项显著性特征及零度层高度的分类模型, 并选用支持向量机担当此任.

3.1 支持向量机分类模型

支持向量机 (SVM) 是根据结构风险最小原理, 在特征空间或特征变换空间寻找一个最优分类超平面, 使得两类样本中离最优分类超平面最近的样本点到超平面的距离最大, 当这个超平面位于特征空间时, 该支持向量机是线性的, 位于特征变换空间时则为非线性的. 例如图 9 最优超平面为 H , 直观地看, 离开超平面 H 越远 (即 d_i 越大) 的样本, 其特征与另一类相差越大, 就越属于这个类型.

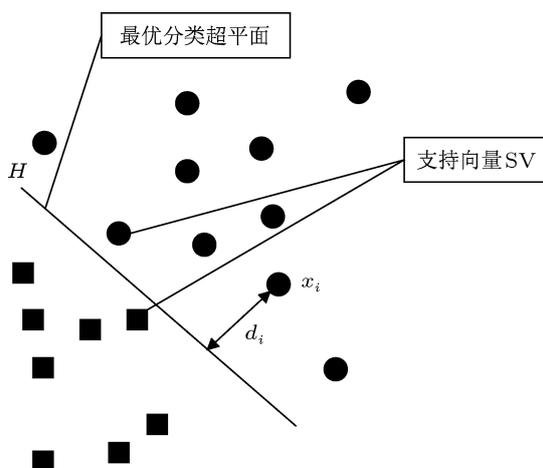


图 9 线性可分 SVM 示意图

设线性可分的样本集 $(x_i, y_i), i = 1, 2, \dots, n$; 其中 x_i 的类别编号 $y_i = \{1, -1\}$, 超平面方程为

$$(\omega \cdot x_i) + b = 0, \quad (18)$$

其中 ω 为权向量, b 为偏置, 在 $\min \frac{1}{2}(\omega \cdot \omega)$ 最小的条件下得到最优判断函数:

$$f(x) = \text{sgn} \left[\sum_{i=1}^n a_i y_i (x_i \cdot x) + b \right] \quad (19)$$

使得

$$y_i [(\omega \cdot x_i) + b] \geq 1, i = 1, 2, \dots, n. \quad (20)$$

在面对非线性分类问题时, 只需将 (19) 式中的点积 $(x_i \cdot x)$ 置换成核函数 $K(x_i, x)$, 即可将其转化为某个隐性的高维空间中的线性问题, 再在该高维空间中得到最优分类超平面^[14]. 在这时的约束条件上添加一个松弛变量 $\xi_i \geq 0$, 使优化问题变为

$$\min \frac{1}{2}(\omega \cdot \omega) + C \sum_{i=1}^n \xi_i, \quad (21)$$

其中惩罚因子 $C > 0$.

非线性支持向量机的最优判断函数为

$$f(x) = \text{sgn} \left[\sum_{i=1}^n a_i y_i K(x_i, x) + b \right]. \quad (22)$$

文献 [15] 研究表明, 利用基于升维思想的支持向量机方法对来自实际气候系统的非平稳过程存在稳定的预测能力. 文献 [16, 17] 研究表明, 结合聚类技术、模糊逻辑的支持向量机在解决混沌时间序列的预测问题时显现优势.

3.2 基于支持向量机的强冰雹识别模型

选用冰雹指数 SHI , 悬垂度 G , 高回波比 ρ_H , 峰度 μ_4 , 有效厚度 ΔH 和零度层高度 H_0 六项特征, 从天津地区 2002—2008 年间 26 次风暴过程的多普勒雷达数据中选出 120 组学习样本 (强冰雹 ($d \geq 20 \text{ mm}$) 60 组, 短时强降水 (降水量 $\geq 20 \text{ mm/h}$) 60 组), 利用 libsvm 库文件编写代码, 训练强冰雹识别模型, 过程如下.

1) 归一化处理: 设特征向量

$$\begin{aligned} X &= [X_1 \ X_2 \ X_3 \ X_4 \ X_5 \ X_6]^T \\ &= [SHI \ G \ \rho_H \ \mu_4 \ \Delta H \ H_0]^T, \end{aligned} \quad (23)$$

对 $X_i, i = 1, 2, \dots, 5$, 使用对各个特征测试过程中获得的 $X_{i\min}$ 和 $X_{i\max}$; 对 X_6 , 根据零度层高度一般处于 2—5 km 之间的经验, 令 $X_{6\min} = 2, X_{6\max} = 5$, 利用 (24) 式, 对样本 $x = (x_1, x_2, \dots, x_6)^T$ 的 6 个分量归一化:

$$x'_i = \begin{cases} 0 & x_i < X_{i\min} \\ \frac{x_i - X_{i\min}}{X_{i\max} - X_{i\min}} & X_{i\min} \leq x_i \leq X_{i\max} \\ 1 & x_i > X_{i\max} \end{cases} \quad (24)$$

2) 确定核函数: 采用最常用的径向基函数作为支持向量机的核函数, 表达式为

$$K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \quad \gamma > 0. \quad (25)$$

3) 训练样本以获得模型参数: 通过对样本的训练, 得到最优惩罚参数 $C = 32$, 最优核参数 $\gamma = 0.5$.

4) 获得样本点到超平面距离 d .

3.3 构建新冰雹指数

通过训练样本得到的由为数不多的支持向量 (样本) 获得的超平面, 将隐性的高维空间一分为二, 强冰雹样本均落入该超平面的一侧, 直观地看, 离开该超平面越远的强冰雹样本与其反例 (短时强降

水) 样本越不像. 换句话说, 远离超平面的样本较超平面附近的样本其强冰雹属性更强. 因此, 文本借助 (26) 式将此距离归一化, 并称之为新冰雹指数:

$$\mu = 2(1 + e^{-Cd})^{-1} - 1, \quad (26)$$

(26) 式中, 参数 C 为决定 μ 变化到 ± 1 快慢的整数, 根据 $2 < d_{\max} < 3$ 的实验结果, 令

$$\mu(3) = 2(1 + e^{-3C})^{-1} - 1 > 0.95, \quad (27)$$

解得 $C > 1.2$, 因此取 $C = 2$.

4 实验测试

4.1 有效性测试

从京津冀地区 2002—2010 年期间 17 个强冰雹过程中随机选取未参加训练的单体样本 33 个、15 个短时强降水过程中选取未参加训练的单体样本 31 个、7 个小冰雹过程中选取单体样本 13 个, 组织对本文提出的新冰雹指数与传统的冰雹指数的对比性测试. 图 10 给出了所有测试样本的传统冰雹指数与零度层高度的散点图, 其中直线 $WT = 57.5H_0 - 121$ 是基于零度层高度的阈值函数; 图 11 是这些样本关于新冰雹指数的分布直方图, 其中横坐标是对新冰雹指数按照四舍五入的方式保留 1 位小数的整定结果. 表 2 是两种方法关于强冰雹击中率 POD 、误报率 FAR 以及临界成功指数 CSI (即 TS 评分) 的统计结果, 其中击中率、误报率及临界成功指数定义如下:

$$POD = \frac{\text{报对的强冰雹单体数}}{\text{强冰雹单体数}}, \quad (28)$$

$$FAR = \frac{\text{报错的强冰雹单体数}}{\text{报对的和报错的强冰雹单体数的和}}, \quad (29)$$

$$CSI = \frac{\text{报对的强冰雹单体数}}{\text{强冰雹单体数} + \text{报错的强冰雹单体数}}, \quad (30)$$

4.2 时效性测试

选择发生于河南、广东的 4 个强冰雹过程的 5 组体扫数据, 组织对本文提出的新冰雹指数与传统冰雹指数的时效性测试, 测试样本的冰雹发生时间等详细信息及测试结果见表 3. 容易看出, 对于 5 组体扫数据所展现的 4 个强冰雹过程, 本文方法全部识别正确, 且识别时间较强冰雹发生时刻至少提

前了 30 min; 而对于相同的测试过程, PUP 所提供的传统冰雹指数仅对河南的强冰雹过程的识别结果与本文相似, 其中对于河南濮阳雷达的探测过程, PUP 的预警提前量比本文算法提前了 2 个体扫 12 min, 至于广东的 3 个过程, PUP 的一个预警提前量大幅度减小、一个仅击中前 11 个体扫且于强冰雹尚未发生时就出现了错误识别、第三个过程的提前识别量从本文方法的 30 min 减小到 6 min.

表 2-1 以短时强降水单体做反例的强冰雹识别

	$POD/\%$	$FAR/\%$	$CSI/\%$
本文	96.97	3.13	91.18
传统	93.93	32.61	60.78

表 2-2 以短时强降水和小冰雹单体做反例的强冰雹识别

	$POD/\%$	$FAR/\%$	$CSI/\%$
本文	96.97	17.95	80.00
传统	93.93	46.55	51.67

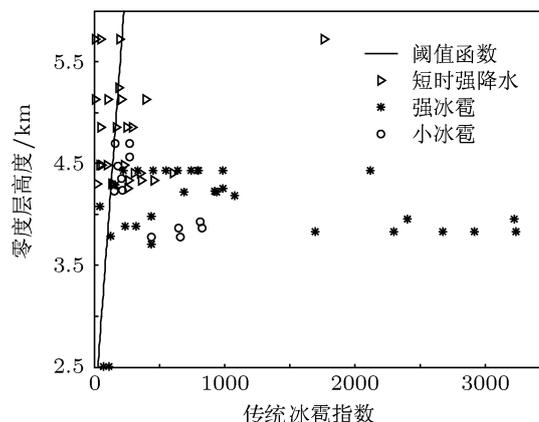


图 10 所有测试样本关于传统冰雹指数与零度层高度的散点图

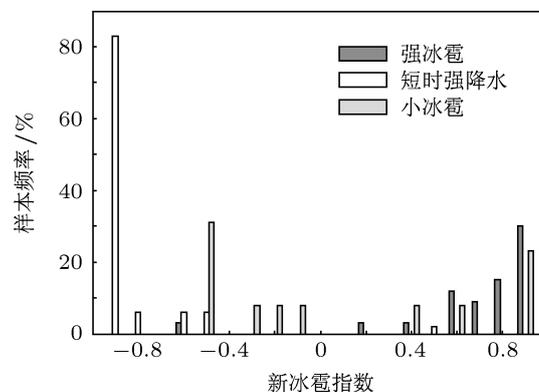


图 11 所有测试样本关于新冰雹指数的分布直方图

表3 时效性测试结果

大冰雹发生日	2008.6.3	2008.6.3	2011.4.17	2012.4.12	2012.4.16
雷达站	河南濮阳	河南商丘	广东广州	广东清远	广东梅州
体扫时段(体扫数)	5:41—7:12 (16)	5:48—9:40 (38)	1:53—6:35 (48)	14:35—16:47(22)	12:23—13:47(14)
冰雹实况	6:44 开始 3—20 mm	6:44 开始 3—20 mm	6:00 后大冰雹	16:47 30 mm	13:00 30 mm
新冰雹指数识别	最初两个体扫没有击中	全部击中	全部击中	全部击中	全部击中
新冰雹指数提前量	54 min	54 min	约 4 h	2 h 18 min	30 min
传统冰雹指数识别	全部击中	全部击中	灾情前 3 个体扫报错	15:47 之后的体扫报错	仅击中 5 个体扫
传统冰雹指数提前量	66 min	54 min	42 min	错误	6 min

5 分析和小结

1) 本文设计了 4 个新的基于风暴回波强度(反射率因子)图像的特征,连同传统的冰雹指数,力图从不同的侧面去贴合强冰雹的物理概念,体现强冰雹单体回波的结构特点,图 8 给出的测试结果表明,就单个特征而言,它们对强冰雹和强降水单体的区分能力均较高,但因在两类间有一定的重叠又不尽善尽美.就传统冰雹指数而言,许多短时强降水由于其强回波顶也会较高(图 8(e),约 7 成的短时强降水的强回波顶越过零度层高度 4 km 以上),因而获得的冰雹指数值也较大,这就是为什么虽然传统强冰雹指数在具有较高的强冰雹击中率(93.93%)的同时,会伴随着很高的误报率(32.61%)的原因.

2) 本文在关注强冰雹单体自身特点的同时,注意到它们与非强冰雹单体量上的差异,令从不同角度抽取的多个显著性特征在模式识别的概念下协同工作,加上支持向量机这种适应性极高的非线性分类器,使得本文归纳出的新冰雹指数呈现出优良的工作效果(表 2-1: $POD = 96.97\%$, $FAR = 3.13\%$, $CSI = 91.18\%$).

3) 因为本文提出的新冰雹指数实质上是从由强冰雹和短时强降水这两类对象的非线性分类器的判断结果转化而来,这个分类器应该同时具有识别短时强降水的能力,站在识别强降水的角度得到统计结果如表 4 所示.可见此模型识别降水的能力依然很强($POD = 96.77\%$, $FAR = 3.23\%$, $CSI = 93.75\%$).

表4 短时强降水识别

	$POD/\%$	$FAR/\%$	$CSI/\%$
以强冰雹单体作反例	96.77	3.23	93.75
以强、弱冰雹单体作反例	96.77	21.05	76.92

4) 因为决定本文新冰雹指数的“悬垂度”、“有效厚度”、“高回波比”等 5 个关键特征于强冰雹的

形成期间就逐渐显著起来,因此新冰雹指数具有识别出正在形成的(即尚未发生的)强冰雹的能力,表 3 给出的测试结果表明,这一提前识别的时效性优于传统的冰雹指数.

5) 按照模式识别的概念,判别模型只适用于对与训练样本相同类型的未知样本进行识别,而强冰雹、短时强降水与弱冰雹乃至弱降水同属对流天气,在线识别时,弱冰雹等是不可回避的识别对象,如果不顾及它们的存在,定会造成在线识别的误报率升高(表 2: FAR 从 3.13% 升至 17.95%; 表 4: FAR 从 3.23% 升至 21.05%).

分析图 11 容易发现,就预报指数而言,有 46.15% 的小冰雹落入了强冰雹区,余下的 53.85% 落入了短时强降水区,这应该是提取特征时没有考虑小冰雹单体特点的缘故.另外,我们注意到小冰雹与短时强降水的指数几乎不发生重叠,对此,本文提出如下解决思路:

1) 利用关键特征(如强回波比)的阈值将弱降水剔除;

2) 增加强降水的特有特征,如液态水含量、时间累积等,将新冰雹指数为负的小冰雹乃至误报为短时强降水的强冰雹从短时强降水中剥离出来;

3) 对弱冰雹和强冰雹展开新特性的设计及其差异性研究,如中气旋及其强度等,寻求从强冰雹中剥离出弱冰雹的有效方法;

4) 添加算法,通过“三体散射”、“强中气旋”等的识别,首先将满足强冰雹充分条件的单体挑选出来;

5) 建立分层识别模型,第一层剔除弱降水,第二层识别满足充分条件的强冰雹,第三层使用本文算法将待识别单体暂时归为强冰雹或短时强降水,第四层进行是否为弱冰雹的鉴别.

以上思路也正是本文工作的进一步的延续,同时本文作者将进一步关注和致力解决冰雹尺寸的预测问题.

- [1] Makitov V 2007 *Atmo. Res.* **83** 380
- [2] Serafin R J, Wilson J W 2000 *Bull. Amer. Meteorol. Soc.* **81** 501
- [3] Holleman I, Wessels H R A, Onvlee J R A, Barlag S J M 2000 *Phys. Chem. Earth B-Hydrol. Oceans Atmos.* **25** 1293
- [4] Lopez L, Sanchez J L 2009 *Atmos. Res.* **93** 358
- [5] Baeck M L, Smith J A 1998 *Weather Forecast.* **13** 416
- [6] Edwards R, Thompson R L 1998 *Weather Forecast.* **13** 277
- [7] Lemon L R 1998 *Weather Forecast.* **13** 327
- [8] Doviak R J, Bringi V, Ryzhkov A, Zahrai A, Zrnic D 2000 *J. Atmos. Ocean. Technol.* **17** 257
- [9] Liang Q Q, Lin L X 2008 *Meteorol. Sci. Technol.* **2** 150 (in Chinese) [梁巧倩, 林良勋 2008 气象科技 **2** 150]
- [10] Zhang G C 2011 *Strong Convection Weather Analysis and Forecasting* (1st Ed.) (Beijing: China Meteorological Press) p186 (in Chinese) [章国材 2011 强对流天气分析与预报 (第一版) (北京: 气象出版社) 第 186 页]
- [11] Zhou Y F, Zhu Y J 2007 *J. Anhui Agricult. Sci.* **35** 9637 (in Chinese) [周叶芳, 朱拥军 2007 安徽农业科学 **35** 9637]
- [12] Witt A, Eilts M D, Stumpf G J, Johnson J T, Mitchell E D, Thomas K W 1998 *Weather Forecast.* **13** 286
- [13] Lenning E, Fuelberg H E, Watson W I 1998 *Weather Forecast.* **13** 1029
- [14] Zhang J S, Dang J L, Li H C 2007 *Acta Phys. Sin.* **56** 67 (in Chinese) [张家树, 党建亮, 李恒超 2007 物理学报 **56** 67]
- [15] Wang F F, Zhang Y R 2012 *Acta Phys. Sin.* **61** 084101 (in Chinese) [王芳芳, 张业荣 2012 物理学报 **61** 084101]
- [16] Cai J W, Hu S S, Tao H F 2007 *Acta Phys. Sin.* **56** 6820 (in Chinese) [蔡俊伟, 胡寿松, 陶洪峰 2007 物理学报 **56** 6820]
- [17] Cui W Z, Zhu C C, Bao W X, Liu J H 2005 *Acta Phys. Sin.* **54** 3009 (in Chinese) [崔万照, 朱长纯, 保文星, 刘君华 2005 物理学报 **54** 3009]

Severe hail identification model based on saliency characteristics*

Wang Ping Pan Yue[†]

(School of Electrical and Automation Engineering, Tianjin University, Tianjin 300072, China)

(Received 31 August 2012; revised manuscript received 13 November 2012)

Abstract

There are always high false alarm ratios when warning against the severe hail with the severe hail index (SHI) which is supplied by digital weather radar system. To solve this problem, the extraction algorithm with several novel features, such as “overhang”, is designed and realized, and these features can describe the severe hail conceptual model from different aspects. Then we take short-time heavy rainfall cells which are easy to be confused with severe hail cells as counter examples to perform statistic analysis for these features and the SHI. Test results show that they have more significant difference between two kinds of samples and hence each of them can reflect one aspect characteristic of severe hail cells. Then a severe hail recognition model that is the Support Vector Machine with radial primary kernel function is learned. Finally, the normalized distance between the sample to be recognized and the optimal separating hyper-plane is determined as a new SHI for warning against the severe hail. Experimental results show that the method proposed in this paper makes severe hail hit ratio higher than the SHI being used and the false alarm ratio is reduced substantially.

Keywords: severe hail recognition, overhang feature of storm cell, significant difference feature, support vector machine

PACS: 92.60.Nv, 92.60.Qx, 92.60.Wc

DOI: 10.7498/aps.62.069202

* Project supported by the Natural Science Foundation of TianJin, China (Grant No. 09JCYBJC07500), and the Special Scientific Research Fund of Meteorological Public Welfare Profession of China (Grant No. GYHY200706004).

[†] Corresponding author. E-mail: panyuetj@gmail.com