

The Extreme Value Machine

Ethan M. Rudd, Lalit P. Jain, Walter J. Scheirer, Senior Member, IEEE, and Terrance E. Boult, Fellow, IEEE

摘要

人们通常希望能够识别在监督下学习的识别函数的输入与培训时未见的类相对应。有了这种能力,人工操作员可以为这些输入分配新的类标签,允许它们被合并到识别功能中——理想情况下是在一个高效的增量更新机制下。尽管存在假定输入来自一组固定类的优秀算法,例如人工神经网络和核心机器,但如何扩展它们以在未知查询类存在的情况下执行增量学习还不是很明显。现有的算法在学习识别函数时很少甚至没有考虑分布信息,而且缺乏强大的理论基础。我们通过构建一个新颖的、理论上可靠的分类器——极值机器 (EVM) 来解决这个问题。EVM 在统计极值理论 (EVT) 的基础上得到了很好的解释,是第一个能够进行非线性无核变带宽增量学习的分类器。与同一深度网络衍生特征空间中的其他分类器相比, EVM 在 ImageNet 数据集的已建立的基准划分上是准确和有效的。

关键词: 机器学习; 监督分类; 开放集合识别; 开放世界识别; 统计极值理论

1 引言

随着时间的推移发展的识别问题需要能够包含新的数据类别的分类器。解决这个问题的方法是什么? 一是定期重新训练分类器。然而,在时间或资源受限的情况下,周期性再训练是不切实际的。另一种可能性是在线类别分类器,它包含了一个有效的增量更新机制。虽然已经提出了解决增量学习问题的方法,但是这些方法的计算成本很高^{[1][2][3][4]},或者几乎没有提供数据统计分布的特征^{[5][6][7][8]},前者是有问题的,因为它违背了使用增量学习的基本动机——高效更新系统的基本动机——而后者限制了推理的质量。

目前的渐进式学习策略也存在一个更根本的问题。当识别系统遇到一个新的类时,这个类应该在随后的增量中被纳入学习过程。但是为了做到这一点,识别系统首先需要识别新的类别。对于这种类型的开集问题,即未知类在查询时出现,我们不能依赖于闭集分类器,即使它支持在线机器学习,因为它隐含地假设所有查询数据都是由训练集很好地表示的。

闭集分类器是为了逼近贝叶斯最优后验概率 $P(C_l | x'; C_1, C_2, \dots, C_M)$, $l \in 1, 2, \dots, M$, 对于一个固定的类集,其中 x' 是输入样本, l 是类 C_l (一个特殊的已知类) 的下标, M 是已知类的数量,在查询时,当出现 Ω 个未知类时,然而,贝叶斯最优后验概率变为 $P(C_{\tilde{l}} | x'; C_1, C_2, \dots, C_M, U_{M+1}, \dots, U_{M+\Omega})$, $\tilde{l} \in \{1, \dots, M + \Omega\}$, 我们无法建立分布模型,因为从 U_{M+1} 到 $U_{M+\Omega}$ 这些类都是未知的。由于来自未知类 $U_{\tilde{l}}$ 的样本 x' 将被错误地归类为已知类 C_l , 因此在训练中进行闭集假设会导致对开集问题的无界支持区域。对于以距离决策边界的符号距离或其校准来评估可信度的分类器来说,如果 x' 距离任何已知数据都很远,这种错误分类就会高度可信地发生——这是一个非常具有误导性的结果。Scheirer 等^[9]将这个问题称为开放空间风险。

更正式地说,设 f 是已知类 C 的可测识别函数, O 是开放空间,因此是一个半径为 r_o 的球,包括所有已知的正训练样本 $x \in C$ 以及开放空间 O , 对于类 C , 其开放空间风险能够被形式化定义为

$R_O(f) = \frac{\int_O f_C(x)dx}{\int_{S_O} f_C(x)dx}$, 其中考虑开放空间风险是正标记开放空间的相对测量值与正标记空间的总体测量值相比。在这种概率公式中, 开集识别的目标是对已知数据无法合理支持的查询行使拒绝选项^[10], 从而减轻这种风险。

开放集合识别^[9], 以及更普遍的新颖性/异常检测都是它们自身已经确立的领域, 但是关于如何在增量环境中处理未知样本的研究却很少, 这是本文的重点。当开放集识别系统将样本标记为未知时, 表明该模型没有用来自对应于该样本的类的数据进行训练。作为回应, 分类器的决策边界应该被更新, 这样系统就可以将这些新的类信息合并到未来的决策中。但是有一个警告: 完整的再培训并不总是可行的, 这取决于时间限制和计算资源的可用性。

最近的工作^[5]扩展了开放集识别问题, 包括在开放世界识别系统中新类的增量学习, 这是我们在在线机器学习中最关心的问题。一个有效的开放世界识别系统必须执行四项任务: 检测未知数, 选择标记哪些点作为模型的补充, 标记这些点, 以及更新模型。一个算法, 最近非离群值 (NNO), 被提出作为这些元素的演示-它的第一个类型。不幸的是, NNO 缺乏强有力的理论基础。该算法使用最近类平均值的阈值距离作为其决策函数, 否则会忽略分布信息。弱分类器是这项工作的一个长期问题: 如何从经典的机器学习理论 (如神经网络和核心机器) 中扩展类边界模型, 使其同时包含增量学习和开集约束, 目前还不是很明显。需要一个新的公式。

在本文中, 我们讨论了基于训练数据分布的开放世界决策边界的紧凑表示的构造。获得这种表示是困难的, 因为在某个时间点上无助于决策边界的训练点可能与以后确定决策边界极为相关, 而在大规模上对所有点进行再训练是不可行的。此外, 由于开放世界问题的定义, 假设空间将被低采样, 因此在许多情况下, 决策边界的线性不能保证, 数据带宽是未知的。那么, 如何在不丢弃潜在相关点的情况下获得一个紧凑的统计模型——特别是在数据带宽是未知的? 为此, 我们引入了极值机 (EVM), 这是一个我们从统计极值理论 (EVT) 派生出来的模型。

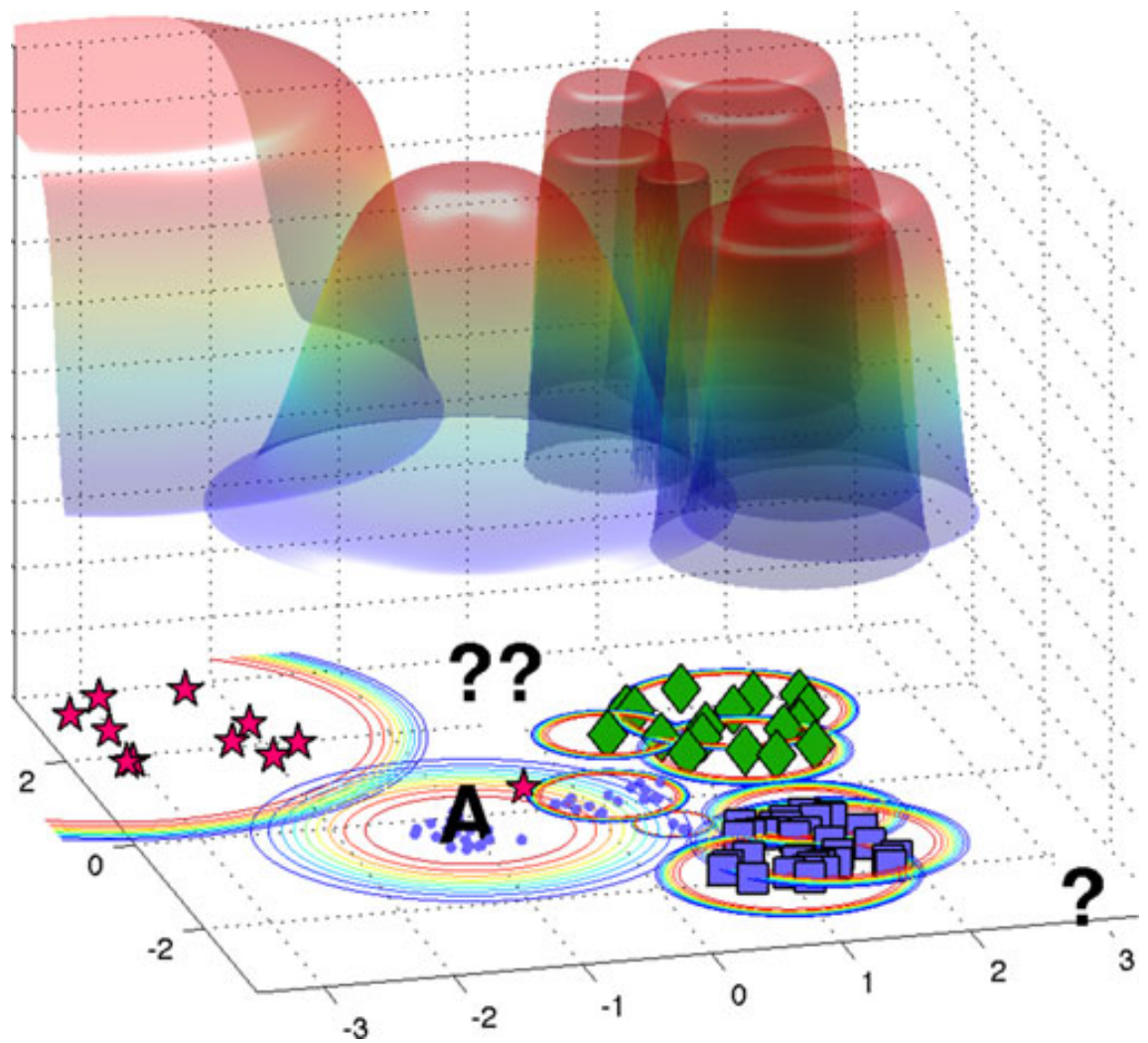


图 1: 一个来自 EVM 算法的解决方案训练了四类: 点, 钻石, 正方形和星星。等高线环上的颜色显示了算法选择的每个极端向量 (EV) 的 Ψ 模型 (样本包含概率), 红色接近 1, 蓝色接近 0.005。通过无核非线性建模, EVM 支持开放集识别, 可以拒绝三个 “?” 超出训练集支持范围的 “未知” 输入每个 Ψ 模型都有自己独立的形状和比例参数, 这些参数都是从数据中学到的, 并且支持软边界。例如, Ψ 模型的蓝点对应的极端向量 A 有一个较为缓慢的下降, 由于在训练期间的异常星的影响。

EVT 规定了一个点相对于另一个点类的径向包含概率的函数形式。通过选择最能概括每个类的点和分布, 即相对于其他类最少冗余的点和分布, 我们得到了每个类的决策边界的一个紧凑的概率表示, 以其极值向量 (EV) 为特征, 它提供了开放空间风险的一个减弱界。这在图 1 中以示意图的形式描述。当新数据到达时, 这些电动汽车可以有效地更新。EVM 是一种可扩展的非线性分类器, 其径向包含函数在某些方面类似于 RBF 核, 但与 RBF 核不同的是, EVT 中的数据导出和接地带有可变的带宽和偏差。

2 相关工作

对于在分类时减少开放空间风险的分类型, 1-vs-Set 机通过用两个超平面包围正数据来代替二元线性分类器的半空间来解决开放集识别问题。针对目标检测问题, Cevikalp 和 Triggs 提出了一种类似于 1-vs-set 机的算法。在后来的工作中, Scheirer^[9] 等人将 W-SVM 引入到使用非线性核的多类开集识别问题中, 并证明了开放空间风险降低的保证。这些非线性耳朵模型更准确, 但计算和存储成本也更高。对于更广泛的开放世界识别问题, Bendale 和 bolt 通过限制模型组合和变换空间的开放空间风险修正了最近类 Mean 算法, 导致了引入的 NNO 算法。

2.1 涉及未知类别数据的相关问题

对于涉及未知类别数据的相关问题，还存在其他方法，如多类别异常检测、域适应和零镜头分类。然而，这些问题不需要通过分类器来解决，分类器随着时间的推移不断更新特定类的特征数据。更相关的是终身学习问题，其中分类器接收任务，并能够调整其模型，以便在新的任务实例上很好地执行。Pentina 和 Ben-David 为基于支持向量机的终身学习任务提出了一个令人信服的理论框架，但是为具体实现这个框架打开了大门。沿着这些思路，Royer 和 Lampert 描述了一种分类器适应的方法，当测试数据中存在固有依赖性时是有效的。这适用于细粒度的识别场景，但不处理在视觉外观上与已知类和其他未知类很好分开的未知类。与我们工作最相关的问题是罕见的类发现，为此，Haines 和 Xiang 提出了一种联合解决类发现和分类任务的主动学习方法。尽管没有区分常见和罕见的未知类。

2.2 统计极值理论

人们对视觉识别的统计极值理论越来越感兴趣。通过观察到任何距离或相似性得分分布的尾部必须始终遵循 EVT，高度精确的概率校准模型成为可能，从而导致多生物特征融合、可描述的视觉属性和视觉检查任务的强有力的经验结果。EVT 模型也被应用于特征点匹配，其中瑞利分布被用于单应估计的有效指导抽样，极值抽样一致性的概念与 RANSAC 被用于类似的手段。机器学习的研究表明，EVT 是一个适合于开集识别问题的模型，其中决策边界的单侧和双侧校准模型导致更好的泛化。然而，这些是在训练时不应用 EVT 的事后方法。

3 本文方法

3.1 理论基础

如第 1 部分所讨论和图 1 所示，EVM 训练集中的每个类都由一组极值向量表示，其中每个向量与一个径向包含映射函数相关联，该函数模拟样本包含概率 (Ψ 或 Ψ)。在这里，我们从 EVT 中推导出 Ψ 的函数形式，这种函数形式不仅在数学上是一个方便的选择，而且在统计上保证了在连续性和平滑性的较小假设下，数据点之间相对接近的极限分布。

本文开发的 EVM 公式源于边际分布的概念。这个想法并不新鲜，人们已经探索了边际分布的各种定义和用法，包括最大化平均或中位数边际、采用加权组合边际或优化边际均值和方差等技术。利用边缘分布本身可以提供比软边缘支持向量机分类器更好的误差界限，这在某些情况下可以转化为减少实验误差。我们模型 Ψ 根据样本半程相对于参考点的分布，从每类公式扩展了边际分布理论公式，该模型适用于每个正参考点的边缘分布 (距离最近的负样本的一半距离)。从这个分布出发，我们推导出一个径向包含映射函数，它刻画出关于参考点的后验概率。这个径向包含映射函数在边缘向零下降。

到目前为止，只有一个标签为 y_i 的类的单个正实例 x_i 。在给定 x_i 的情况下，最大的边际距离是距离最近的训练样本距离的一半。然而，最接近的点只是一个样本，我们应该考虑在不同样本下的潜在最大边缘。我们将边际分布定义为观测数据的边际距离的分布。因此，给定 x_i 和 x_j ，其中 $\forall j, y_j \neq y_i$ ，考虑到对 (x_i, x_j) 估计的决策边界的边界距离 τ 如果 x_j 是最近的实例。因此，对于 τ 个最接近点，边界估计是 $m_{ij} = \|x_i - x_j\| / 2$ 。考虑到这些最接近边界的 τ 点，我们的问题就变成了：边界距离的分布形式是什么？

为了估计这个分布，我们求助于 Fisher-Tippett 定理，也被称为极值定理。正如中心极限定理指出，

从某些随机过程产生的随机变量遵循高斯分布，EVT 指出，给定一个表现良好的整体分布的值，例如，一个分布是连续的，有一个逆，最大值或最小值的分布只能采取有限的形式。为了找到合适的形式，极大值理论：

Theorem 1 (Fisher-Tippett Theorem [37]). Let $\zeta_n = \max \{v_1, \dots, v_n\}$

$$\lim_{z \rightarrow \infty} P \left(\frac{\zeta_n - b_n}{a_n} \leq z \right) = F(z)$$

参考：对于中心极限定理来说，无论一个随机变量 X 服从什么分布，每次取样一批随机变量的均值 (或和) 都服从正态分布。即

$$\lim_{n \rightarrow \infty} F_n(x) = \lim_{n \rightarrow \infty} \left\{ \frac{\sum_{s=1}^n X_i - n\mu}{\sqrt{n}\sigma} \leq x \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt = \Phi(x)$$

再来考虑这个最小半距离 mij，应该是最小值分布，因此，z=-mij 则是最大值分布（很接近上面的极值分布了），由于 z=-mij<0, 所以 z=-mij 服从逆向极值分布，而 mij=-z 则服从的分布叫做韦伯分布，它指的是，无论一个随机变量 X 服从什么分布，每次取样一批随机变量的最小值都服从韦伯分布。

$$F(x) = \begin{cases} 1 - \exp[-(x/\eta)^m], & x > 0 \\ 0, & x \leq 0 \end{cases}$$

所以我们可以认为这个最小半距离服从韦伯分布，即这个最小半距离（边界距离）mij<xi 的概率为 F(xi)。对于任意一个点 x'，当且仅当它到这个类中心 x 的距离 ||x'-xi|| 小于这个最小半距离时，这个点 x' 才会属于这个类，因此一个点 x' 属于这一类的概率则可以表示为 1-F(||x'-xi||) 即得到相应模型：

$$\Psi(x_i, x'; \kappa_i, \lambda_i) = \exp - \left(\frac{\|x_i - x'\|}{\lambda_i} \right)^{\kappa_i}$$

上面说到样本点 x' 属于 Ci 这一类的概率为 1-F(||x'-xi||), 那么 x' 属于其他已知类 C1,C2... 的概率则是 1-F(||x'-x1||), 1-F(||x'-x2||) ..., 在闭集多分类中，我们首先找到最大概率类 Ci, 使得

$$\hat{P}(C_l | x') = \max_{\{i: y_i = C_l\}} \Psi(x_i, x'; \kappa_i, \lambda_i)$$

然后如果这个最大概率大于给定阈值的话（这个给定阈值的选择主要是基于已知类识别的准确率和未知类识别/拒绝的准确的平衡），则多分类的结果为 Ci, 否则分类为未知类，即使用下面的决策公式

$$y^* = \begin{cases} \arg \max_{l \in \{1, \dots, M\}} \hat{P}(C_l | x') & \text{if } \hat{P}(C_l | x') \geq \delta \\ \text{"unknown"} & \text{Otherwise.} \end{cases}$$

3.2 Model Reduction

当训练集中每一类的样本规模太大时，为了避免在测试中，计算与每一个类中样本的最近距离时，需要考虑的点太多，需要将代表一个类的点减少，即首先为每一个类，选择最小的表征点数。选择的思路在于，如果一个类中 xj 与 xi 样本的“距离”越小（属于同一类的概率 1-F(||xj-xi||) 越大, 大于给定阈值）时，则可以删掉这个点。而选择最下的表征点数在删除这些点时，还要考虑点和点之间的关

系，即要保持最终留下来的点是最少的，这样留下来的点称为极值向量（EVs）。即实现指示函数：

$$I(x_i) = \begin{cases} 1 & \text{if } \langle x_i, \Psi(x_i, x', \kappa_i, \lambda_i) \rangle \text{ kept} \\ 0 & \text{otherwise.} \end{cases}$$

这个问题属于集合覆盖问题，是 NP 难问题，于是使用了贪婪算法来做，可以在多项式内求解。除了通过阈值进行限制点个数之外，还可以直接限制样本点的个数。

$$\begin{aligned} &\text{minimize } \sum_{i=1}^{N_l} I(x_i) \text{ subject to} \\ &\forall j \exists i \mid I(x_i) \Psi(x_i, x_j, \kappa_i, \lambda_i) \geq \varsigma. \end{aligned}$$

注意，在模型拟合中，一个异常点通常被另一个类的一个点所覆盖（见图 1），而且这样的异常点也不太可能覆盖许多其他的点。因此，异常值被添加到覆盖集非常晚，如果有的话。这不是一个特别的假设；它是覆盖所有例子的点数最小化过程的结果。就像边缘固有的柔软性一样，这是 evt 建模之后的模型简化方法的固有部分。

3.3 增量学习

一旦 EVs 已经为 EVM 获得，模型可以通过使用当前 EVs 和新批数据中的所有点来拟合新数据的 Ψ 模型来更新新批数据。新的极值向量是通过在旧的极值向量和新的训练点上运行模型缩减得到的。虽然可以单独添加新点，但批量添加数据将产生更有意义的匹配，因为批量在每次增量时代表更丰富的分布样本。这意味着新增加的训练点可能成为 EVs，也可能不成为 EVs，新的类也可能影响以前学习的模型和 EVs。在有效的模型简化技术允许 EVM 通过显式选择 ς 或者在最大 k 覆盖贪婪方法中通过特定的最大 EVs 数限制其模型的大小。这允许 EVM 通过不同的操作模式扩展到许多不同的增量问题。例如，我们为添加数据的类在每个训练增量处选择一个静态 ς 并使用该阈值执行模型缩减。虽然这限制了模型大小，极值向量的数量仍然随着时间的推移而增加。有界优化将指定每个类的最大大小或最大总大小，重新计算 ς 在每个增量。或者，最大模型大小可以通过模型缩减预先指定，只有当最大模型大小被违反时才执行。因此，EVM 不仅是一个增量分类器，而且是一个增量分类器，其大小可以在每个学习增量处进行控制。

增量学习不仅仅可以针对未知类，还可以针对已知类增加训练样本。在增针对未知类时，就可以直接进行一次对这个类的最小表征点选择过程。而针对已知类增加训练样本，则可以完全不管之前已经被抛弃的点，直接利用新加入的点和此前的极值向量（EVS）进行一次最小表征点选择过程。

4 复现细节

4.1 与已有开源代码对比

参考代码 <https://github.com/EMRRResearch/ExtremeValueMachine> 部分代码，其中参考其 evm 算法代码中相关的加载数据、威布尔拟合部分。

相关伪代码描述：

Procedure 1 set cover greedy.

Input: the number of C_i Class :U; subsets from set over :subsets.

Output: kept EVs of C_i Class :cover_indices.

$covered = \phi$

while The number of covered points is not equal to u **do**

$max_index = \text{Max}(\text{len}(x - covered))$ for x in subsets)

 ▷Take the point with the largest coverage

 Add subsets[max_index] to $covered$

 ▷add the point and the coverage of the point to the covered

 Add (max_index) to $cover_indices$

 ▷add the point to the $cover_indices$

end

Procedure 2 set cover.

Input: points of C_i class:points; weibulls of these points :weibulls; set_cover_greedy :S; Euclidean distance function :D; cover_threshold :K.

Output: kept EVs of C_i class :keep_indices.

U=len(points)

▷Return the number of points

d_mat = D(points)

▷Return the distance matrix between similar points

Use weibulls to Convert d_mat to Probability get probs

Find all subsets with probability greater than K get subsets

keep_indices=S(U,subsets)

▷Return to remaining points

Procedure 3 reduce model.

Input: points; weibulls; labels; cover_threshold :K; set_cover :S; labels_to_reduce.

Output: kepted points :keep_points; kepted weibulls keep_weibulls; kepted labels :keep_labels.

if $K \geq 1$ **then**

 return points, weibulls, labels;

 ▷do not to reduce model

else

for ulabel **in** labels_to_reduce **do**

 Find out all points of the ulabel And weibulls of corresponding points ind_points, ind_weibulls

 keep_ind=S(ind_points, ind_weibulls)

 ▷Return to remaining points

 add keep_ind to keep_points

 add weibulls of corresponding keep_ind to keep_weibulls

 add labels of corresponding keep_ind to keep_labels

end

end

4.2 实验环境搭建

Requirements: 实验所需的环境如下

- python > 3.6
- libMR
- pandas
- numpy
- sklearn
- matplotlib
- seaborn

- hyperopt

需要 python3.6 版本以及以上的。

需要安装 libMR、pandas、numpy、sklearn、matplotlib、seaborn、hyperopt 这些库。

Installation:

进行命令: `pip install libmr, pandas, numpy, sklearn, matplotlib, seaborn, hyperopt`

代码结构介绍:

- 'code_with_evm' 包含 EVM 相关代码的文件夹。
 - config.py 包含 EVM 参数配置信息。
 - Code_of_EVM.py 包含 EVM 算法代码。
 - Score_of_EVM.py 包含评估信息, Recognition Accuracy:F-Measure、Precision、Recall。
 - Turn_Hyperparameter.py 调参数, 对于参数: tailsize、cover_threshold、num_to_fuse、margin_scale、ot 在相应的 (30,200,5)、(0.2,0.95,0.1)、(1,10,1)、(0.3,0.7,0.1)、(0,0.3,0.001) 范围内进行调整, 能够得出最优的参数。
 - test.py 测试代码, 进行代码的测试。
- 'images' 存储相关实验结果图
 - equence.jpg 在 letter 数据集中, W-SVM、rest_svm、NN+CAP、以及 EVM 的 F1-Measure 的对比试验结果图
 - EVM_confusion_matrix.png 存储 EVM 的混淆矩阵图
- 'letter' 关于 letter 数据集相关实验代码
 - equence_plt.py 绘制 letter 数据集中, W-SVM、rest_svm、NN+CAP、以及 EVM 的 F1-Measure 的结果图
 - EVM_of_letter.csv 存储 EVM 在 letter 数据集的相关数据
 - letter-recognition.csv letter 数据集
 - experiment-of-letter-unseen.py 在 letter 数据集中进行 evm 实验
 - others_of_lettter.csv 存储在 letter 数据集中其他对比实验方法得出的数据结果
 - sample_data_prep.py 划分 letter 数据集
 - test.py 测试代码
- 'letter_data' 存储 letter 训练和测试数据
 - test.csv 在 letter 测试数据集
 - train.csv 在 letter 训练数据集

4.3 界面分析与使用说明

代码结构介绍:

- 'code_with_evm' 包含 EVM 相关代码的文件夹。
 - config.py 包含 EVM 参数配置信息。
 - Code_of_EVM.py 包含 EVM 算法代码。
 - Score_of_EVM.py 包含评估信息, Recognition Accuracy:F-Measure、Precision、Recall。

- Turn_Hyperparameter.py 调参数，对于参数：tailsize、cover_threshold、num_to_fuse、margin_scale、ot 在相应的 (30,200,5)、(0.2,0.95,0.1)、(1,10,1)、(0.3,0.7,0.1)、(0,0.3,0.001) 范围内进行调整，能够得出最优的参数。
- test.py 测试代码，进行代码的测试。
- 'images' 存储相关实验结果图
 - equence.jpg 在 letter 数据集中，W-SVM、rest_svm、NN+CAP、以及 EVM 的 F1-Measure 的对比试验结果图
 - EVM_confusion_matrix.png 存储 EVM 的混淆矩阵图
- 'letter' 关于 letter 数据集相关实验代码
 - equence_plt.py 绘制 letter 数据集中，W-SVM、rest_svm、NN+CAP、以及 EVM 的 F1-Measure 的结果图
 - EVM_of_letter.csv 存储 EVM 在 letter 数据集的相关数据
 - letter-recognition.csv letter 数据集
 - experiment-of-letter-unseen.py 在 letter 数据集中进行 evm 实验
 - others_of_lettter.csv 存储在 letter 数据集中其他对比实验方法得出的数据结果
 - sample_data_prep.py 划分 letter 数据集
 - test.py 测试代码
- 'letter_data' 存储 letter 训练和测试数据
 - test.csv 在 letter 测试数据集
 - train.csv 在 letter 训练数据集

首先是运行 sample_data_prep.py 代码，进行数据分割，再用 Code_of_EVM.py 代码进行其预测，可以使用上述说明的相关代码进行性能评估，也就是分数获取代码，最后用上述的绘图代码进行实验结果的呈现。

4.4 创新点

(1) 威布尔拟合确保了一小部分标签错误的点或其他离群值不会导致从威布尔估计的边界在该位置。如果拟合包括更多的距离点， Ψ 模型将在尺度/形状上扩展，为其边界定义中的“柔性”提供一个自然派生的理论。

(2) 然而，Set Cover 的整体优化目前缺乏一个参数来调整正面和负面类别之间的风险权衡。未来的研究方向可能包括通过在 Set Cover 期间获得更好的参数化软边界直接扩展 EVM，也许可以通过增加权值来平衡软边界误差，并用线性规划的方法来表述问题。

(3) 另一个潜在的扩展是在一个支持向量机风格的优化算法的损失函数中合并边际权重。

5 实验结果分析

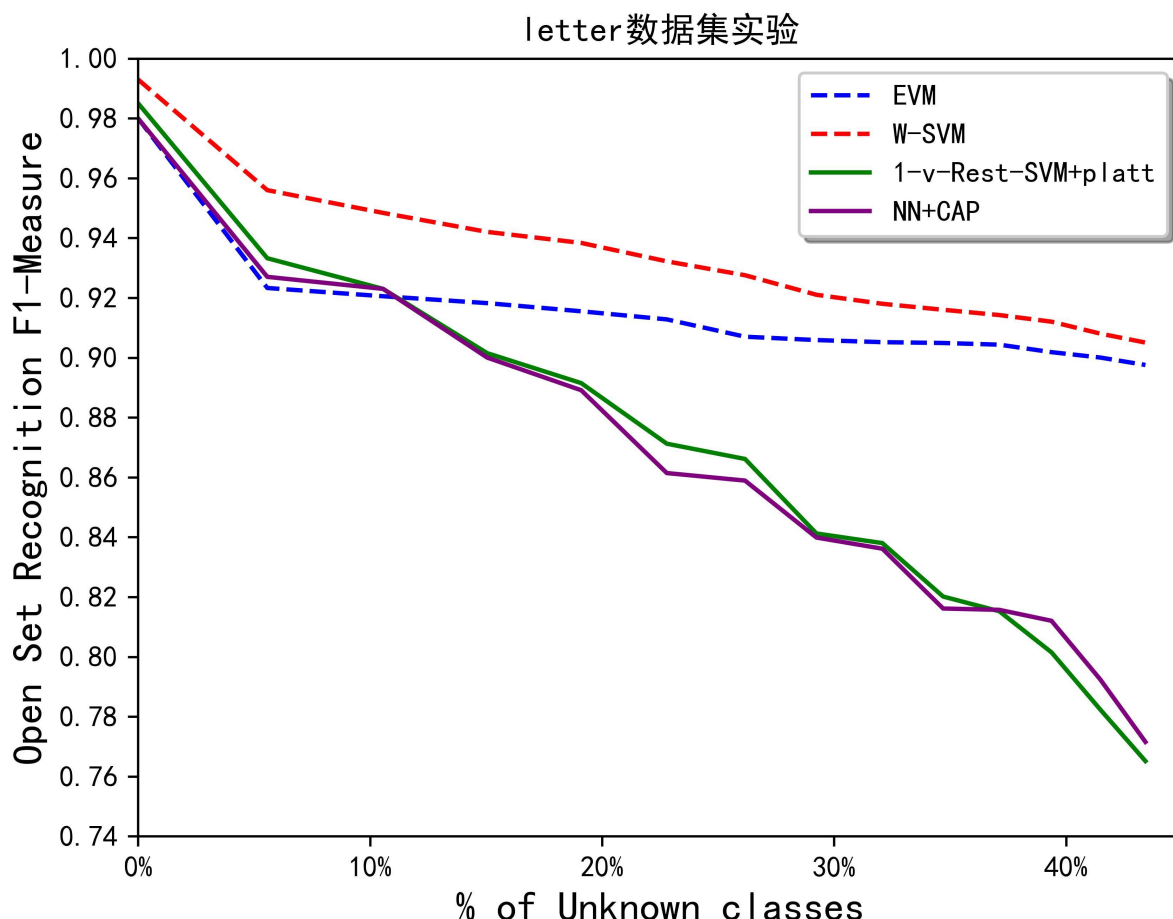


图 2: 实验结果示意

图 2 显示了所有评估算法的结果，包括开集特定 w-svm 支持向量机，这是目前文献中对这个问题最好的算法。给出了具有 CAP 概率估计的最近邻 (NN) 分类器和具有 Platt 概率估计的 1-vs-rest RBF 支持向量机的结果。评估的其他校准模型的表现显着更差，并且没有显示。在训练期间从 Letter 数据集中随机选择 15 个不同的标签作为已知类，并通过在测试期间逐步包含剩余 11 个标签的子集来添加未知类。这个过程重复超过 20 倍以计算平均值和误差线。用 F1-Measure (过度准确率召回率) 来报告结果。

EVM 算法的性能与 w-svm 支持向量机相当，优于其他所有算法。w-svm 支持向量机无疑是一种可行的数据集算法，但它的微小优点比 EVM 的成本更高，需要两个训练有素的支持向量机模型 (每个类: 一个 1 类和一个二进制) 进行操作。这个实验的矢量比是针对所有 26 个类的模型进行计算的。对于图 2 中的评估，EVM 的向量比值比任何支持向量机模型的向量比值小一个数量级，表明对于选择的阈值 (0.5)，不到一半的训练数据点包含在 EVs 中。SVM 模型中支持向量的数量大于训练集中的点数。

O-LETTER 上的多类开集识别性能。X 轴代表测试集中在训练期间看不到的类的百分比 (未知类)。EVM 与 f1 测量中现有的最先进的 W-SVM 相当，但在训练效率和可伸缩性方面有很大的节省，如向量比 (VR) 所反映的。EVM 的 VR 比两个基于 svm 的模型小一个数量级。当问题变得更加开放时，EVM 和 W-SVM 算法都具有良好的性能退化特性。另外两个经过概率校准的算法的退化速度要快得多。采用 75(尾部大小) 和 4(平均超过的 EVs 数) 的超参数进行评价，并采用与 w-svm 相同的训练集交叉验

证技术进行选择。

虽然通过缓存重复的支持向量可以降低 platt 标定的 SVM 模型的处理和存储成本,但由于 w-SVM 采用不同的多核支持向量机模型,因此计算量的节省是不可行的。这是因为不同的 RBF 核需要不同的计算,即使它们集中在同一点上。此外,对于 EVM,不像 W-SVM,我们可以很容易地得到一个较低的矢量比,同时最小化任何降低的精度。最后,提到的是,除了 EVM 之外,图 3 中描述的性能的分类器都不支持增量学习,因此它们不能应用于开放的世界问题。

6 总结与展望

这项工作最重要的结论是 EVM 能够进行无核非线性分类。有趣的是, EVM 与径向基函数有一些共同的关系。当 $k=2$ 时,与高斯 RBF 相同,当 $k=1$ 与指数 RBF 或拉普拉斯 RBF 相同。虽然这些 k 值在实践中可能出现,但 k 假设的值范围更广,通常更大。此外,对于 $k > 2$,不是 Mercer 内核。通过高斯 (或拉普拉斯) 的加权和我们有两种不同的方式来看待一个高斯 (或拉普拉斯) RBF 核的近似 Ψ 模型。当 Ψ 模型参数的尺度和形状随着数据集的带宽和密度的变化而变化时,在高斯近似下,核元的数目和或近似的精度必须在空间上变化。EVM 对于边缘分布及其 Ψ 模型需要最少的点。对于 EVM,我们既不对一个特定的核函数进行特别的假设,也不对一个特定的核函数进行特别的假设, Ψ 模型的函数形式是 EVT 用于模拟输入空间距离分布的直接结果。

威布尔拟合确保了一小部分标签错误的点或其他离群值不会导致从威布尔估计的边界在该位置。如果拟合包括更多的距离点, Ψ 模型将在尺度/形状上扩展,为其边界定义中的“柔性”提供一个自然派生的理论。然而, Set Cover 的整体优化目前缺乏一个参数来调整正面和负面类别之间的风险权衡。未来的研究方向可能包括通过在 Set Cover 期间获得更好的参数化软边界直接扩展 EVM,也许可以通过增加权重来平衡软边界误差,并用线性规划的方法来表述问题。另一个潜在的扩展是在一个支持向量机风格的优化算法的损失函数中合并边际权重。

参考文献

- [1] CRAMMER K, DEKEL O, KESHET J, et al. Online passive aggressive algorithms[J]., 2006.
- [2] YEH T, DARRELL T. Dynamic visual category learning[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition. 2008: 1-8.
- [3] LASKOV P, GEHL C, KRÜGER S, et al. Incremental support vector learning: Analysis, implementation and applications.[J]. Journal of machine learning research, 2006, 7(9).
- [4] LI L J, FEI-FEI L. Optimol: automatic online picture collection via incremental model learning[J]. International journal of computer vision, 2010, 88(2): 147-168.
- [5] BENDALE A, BOULT T. Towards open world recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1893-1902.
- [6] KAPOOR A, BAKER S, BASU S, et al. Memory constrained face recognition[C]//2012 IEEE Conference on Computer Vision and Pattern Recognition. 2012: 2539-2546.

- [7] MENSINK T, VERBEEK J, PERRONNIN F, et al. Metric learning for large scale image classification: Generalizing to new classes at near-zero cost[C]//European Conference on Computer Vision. 2012: 488-501.
- [8] RISTIN M, GUILLAUMIN M, GALL J, et al. Incremental learning of ncm forests for large-scale image classification[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 3654-3661.
- [9] SCHEIRER W, ROCHA A, SAPKOTA A, et al. Towards Open Set Recognition, TPAMI[J]. Cited on, 2013: 54.
- [10] BISHOP C M, NASRABADI N M. Pattern recognition and machine learning: vol. 4[M]. Springer, 2006.