

一种半监督协同训练的正则化算法

王 娇¹ 罗四维²

(中央广播电视大学计算机科学与技术系 北京 100039)¹

(北京交通大学计算机应用研究所 北京 100044)²

摘要 半监督学习是机器学习领域的研究热点。协同训练研究数据有多个特征集时的半监督学习问题。从正则化角度研究协同训练,利用假设空间的度量结构定义学习函数的光滑性和一致性,在每个视图内的学习过程中以函数光滑性为约束条件,在多个视图的协同学习过程中以函数一致性为约束条件,创新性地提出一种两个层次的正则化算法,同时使用函数的光滑性和一致性进行正则化。实验表明,该算法较仅使用光滑性或仅使用一致性的正则化方法在预测性能上有显著提高。

关键词 人工智能,机器学习,半监督学习,多视图学习,正则化算法

中图分类号 TP181 **文献标识码** A

New Regularization Method of Co-training

WANG Jiao¹ LUO Si-wei²

(Department of Computer Science, The Open University of China, Beijing 100039, China)¹

(Department of Computer Science & Technology, Beijing Jiaotong University, Beijing 100044, China)²

Abstract Semi-supervised learning is a hot research topic of machine learning. Co-training is a multi-view semi-supervised learning method. Co-training is studied from the regularization point of view. By exploiting the metric structure of the hypotheses space, the smoothness and consistency of a hypothesis were defined. A two levels regularization algorithm was presented which uses the smoothness to regularize the within-view learning process and uses the consistency to regularize the between-view learning process. The experimental results were presented on both synthetic and real world datasets.

Keywords Artificial intelligence, Machine learning, Semi-supervised learning, Multi-view learning, Regularization

1 引言

监督学习通过对有标记数据进行学习来构建预测模型。这里,标记是指数据所对应的输出。当有标记数据的数量较少时,监督学习所训练出的学习器往往很难有好的泛化性能。随着信息时代的到来,数据无处不在,但获取数据的标记相对困难。半监督学习正是为了解决这一问题而产生的,它利用大量存在的未标记数据辅助学习,以提高传统监督学习的预测性能,是机器学习领域中近年来国内外研究的热点^[1,2]。

在机器学习的一些实际问题中,数据有多个特征集,也称为多个视图。Blum和Mitchell^[3]在研究网页分类问题时,将网页本身包含的信息作为一个视图,将超链接所包含的信息作为另一个视图,从而用两个视图进行半监督学习,获得了较好的效果,其算法被称为协同训练算法。多视图的半监督学习由此受到研究者的关注^[4,5]。

从有限观测数据中学习的问题常常是不适定的逆问题,正则化是为获得稳定的解而根据先验知识对学习过程加以限制的过程。经典的机器学习理论^[6]通过控制假设空间的复杂

度来提高学习器的预测性能。半监督学习中标记数据非常稀少。半监督学习问题往往是一个不适定的逆问题,需要通过正则化的方法对学习过程加以限制。本文从正则化的角度研究多视图的半监督学习,创新性地提出一种针对多视图半监督学习的正则化算法,以提高学习器的预测性能。提出的正则化算法从光滑性和一致性两个方面来度量学习函数,在每个视图内的学习过程中约束函数的光滑性,在多个视图间的协同学习过程中约束函数的一致性,在真实数据集上验证算法的有效性。

2 一种针对半监督协同训练的正则化算法

在典型的半监督学习中,数据集由两部分组成,即 l 个标记数据组成的集合 $\{(x_1, y_1), \dots, (x_l, y_l)\}$ 和 u 个未标记数据组成的集合 $\{x_{l+1}, \dots, x_{l+u}\}$,其中 $x_i \in X$ 表示数据点, $y_i \in Y$ 表示对应的标记。用带下标的 x_i 表示第 i 个数据点,用带上标的 $x^{(v)}$ 表示数据的第 i 个视图,将多视图数据表示为 $x = (x^{(1)}, x^{(2)}, \dots, x^{(v)})$,其中 v 表示视图的个数。

在多视图的半监督学习中,一方面数据有多个视图,即

到稿日期:2011-10-06 返修日期:2011-12-03 本文受国家自然科学基金项目(60975078)资助。

王 娇(1982—),女,博士,讲师,主要研究方向为机器学习、模式识别, E-mail: wang0828@gmail.com; 罗四维(1943—),男,博士,教授,主要研究方向为神经计算。

$x = (x^{(1)}, \dots, x^{(n)})$; 另一方面数据由标记数据集 $\{(x_1, y_1), \dots, (x_l, y_l)\}$ 和未标记数据集 $\{x_{l+1}, \dots, x_{l+u}\}$ 组成。学习算法应考虑如何利用多个视图蕴含的信息和未标记数据蕴含的信息, 来辅助传统的监督学习。下面首先引入假设空间中的度量结构, 然后使用这个度量定义函数的光滑性和一致性, 将其作为函数的评价指标来指导学习过程, 并由此提出一种针对多视图半监督学习的正则化算法。

2.1 假设空间的度量结构

在监督学习中, 需要学习一个从输入到输出的映射(或称函数) $f: X \rightarrow Y$, 以对真实分布 $P_{X \times Y}$ 进行建模。在半监督学习中, 研究如何利用边缘分布 P_X 这个额外信息来学习更优的函数。

Schuermans 和 Southey^[7] 利用边缘分布信息 P_X 在假设空间中定义了一个度量, 通过计算任意两个函数 f, g 对数据预测的不同, 得到它们之间距离的一种计算方式, 即

$$d(f, g) = \varphi \left(\int \text{err}(f(x), g(x)) dP_X \right) \quad (1)$$

式中, $\text{err}(f, g)$ 衡量函数 f 和 g 预测误差的不同, φ 是使距离 d 满足度量性质的标准化函数。

对于回归问题, 预测误差可以用平方误差 $\text{err}(f, g) = (f - g)^2$ 或其它类似的代价函数来衡量。对于分类问题, 预测误差可以用误分代价函数 $\text{err}(y, \hat{y}) = 1_{(y \neq \hat{y})}$ 来衡量。标准的度量函数需要满足非负性、对称性和三角不等式, 大多数典型的预测误差函数都满足度量函数的要求。例如, 在回归问题中, 两个函数的距离可以表示为

$$d(f, g) = \left(\int (f(x) - g(x))^2 dP_X \right)^{1/2} \quad (2)$$

式中, 标准化函数 $\varphi(z) = z^{1/2}$ 使距离 d 满足度量性质。在分类问题中, 两个分类器的距离可以表示为

$$d(f, g) = \int 1_{(f(x) \neq g(x))} dP_X = P_X(f(x) \neq g(x)) \quad (3)$$

这里不需要标准化就能使距离 d 满足度量性质。

可以推广这种定义方式, 使其包含真实分布。

$$d(P_{Y|X}, h) = \varphi \left(\int \int \text{err}(h(x), y) dP_{Y|X} dP_X \right) \quad (4)$$

即, 函数 h 相对于真实分布 $P_{Y|X}$ 的误差可以用它们之间的距离 d 来表示。这种定义方式与式(1)中的定义方式相同, 它使我们能够在一般的度量空间结构中研究机器学习问题。

回归问题中, 函数 h 与真实分布的距离可以表示为

$$d(P_{Y|X}, h) = \left(\int \int (h(x) - y)^2 dP_{Y|X} dP_X \right)^{1/2} \quad (5)$$

分类问题中, 函数 h 与真实分布的距离可以表示为

$$\begin{aligned} d(P_{Y|X}, h) &= \int \int 1_{(h(x) \neq y)} dP_{Y|X} dP_X \\ &= P_{XY}(h(x) \neq y) \end{aligned} \quad (6)$$

2.2 函数光滑性的定义

我们认为, 一个好的学习器在标记和未标记数据上应该有相似的预测性能, 即不仅在标记数据上的误差较小, 而且要在未标记数据上的误差较小。因为如果一个学习器仅在标记数据上的预测误差较小, 而在未标记数据上的预测误差较大, 则其不可能具有好的泛化性能。

在假设空间 H 中, 用 $d(f, g)$ 表示函数 f 和函数 g 之间的距离。具体地, 用 $d_L(f, g)$ 表示在标记数据集 L 上的距离, 用 $d_U(f, g)$ 表示在未标记数据集 U 上的距离, 用 $d_X(f, g)$ 表

示在整个数据集 X 上的距离, 其分别定义为

$$d_L(f, g) = \varphi \left(\frac{1}{l} \sum_{i=1}^l \text{err}(f(x_i), g(x_i)) \right) \quad (7)$$

$$d_U(f, g) = \varphi \left(\frac{1}{u} \sum_{i=l+1}^{l+u} \text{err}(f(x_i), g(x_i)) \right) \quad (8)$$

$$d_X(f, g) = \varphi \left(\frac{1}{l+u} \sum_{i=1}^{l+u} \text{err}(f(x_i), g(x_i)) \right) \quad (9)$$

给定辅助光滑函数 ϕ , 分别计算函数 f 和光滑函数 ϕ 在标记和未标记数据上的距离, 即 $d_L(f, \phi)$ 和 $d_U(f, \phi)$; 对于函数 g , 同样计算 $d_L(g, \phi)$ 和 $d_U(g, \phi)$ 。如果有

$$|d_L(f, \phi) - d_U(f, \phi)| < |d_L(g, \phi) - d_U(g, \phi)| \quad (10)$$

那么, 称函数 f 比函数 g 光滑。式(10)表示, 与函数 g 相比, 函数 f 与已知光滑函数 ϕ 的距离在标记和未标记数据上相近。即, 与函数 g 相比, 函数 f 在标记和未标记数据上具有相似的预测性能。由此定义函数的光滑性如下。

定义 1 若函数在标记和未标记数据上的预测性能相似, 则其光滑性能好。

由定义 1 得出, 在假设空间 H 中函数光滑性的评价指标为

$$\min_{h \in H} (d_L(h, \phi) - d_U(h, \phi))^2 \quad (11)$$

即, 借助辅助光滑函数 ϕ 寻找最优函数 h , 使它与 ϕ 在标记和未标记数据上的距离相近, 从而使它在标记和未标记数据上的预测性能相似。 ϕ 选择最简单的光滑函数即可, 例如, $\phi = 0$ 或 $\phi = \frac{1}{l} \sum_{i=1}^l y_i$ 。

与传统正则化方法中定义的光滑性相比, 式(11)中定义的光滑性是自适应于数据的。传统正则化方法对任何数据集都使用相同的约束项; 而本文提出的正则化方法针对具体数据, 充分考虑了未标记数据的作用。由式(11)可以看出, 对于不同的数据集, $d_L(h, \phi)$ 和 $d_U(h, \phi)$ 不同, 从而使此光滑性评价指标自适应于具体数据集。

2.3 函数一致性的定义

我们认为, 当使用多个学习器对某数据的输出进行预测时, 这些预测应该具有一致性。所谓一致性是指这些预测结果相似, 在分类问题中代表分为同一类, 在回归问题中代表函数值相同。这样考虑的出发点是, 如果多个学习器对此数据的预测一致, 那么这多个学习器同时出错的可能性较低, 所以它们的预测结果就更可靠。

Sa^[8] 在研究词义消歧问题时, 最早使用了多个独立函数的一致性来降低它们的错误率。Dasgupta 等人^[9] 研究了两个独立函数的一致性与它们的错误率之间的关系, 结论是: 若每个函数的错误率都小于 $1/2$, 则不等式(12)成立。

$$p(h^{(1)}(x) \neq h^{(2)}(x)) \geq \max_{i=1,2} p(h^{(i)}(x) \neq y) \quad (12)$$

式中, $h^{(1)}$ 和 $h^{(2)}$ 代表两个独立函数, y 代表数据 x 的真实标记。式(12)表示, 两个函数不一致的概率是每个视图错误率的上界。

这个结论提供了一种机制, 使我们可以利用函数在未标记数据上的一致性来度量函数的错误率。为此, 定义最大化一致原则如下。

定义 2 当使用多个函数对数据进行决策时, 最大化这多个函数的一致性。

使用 2.1 节中的度量结构, 最大化函数 $h^{(1)}$ 和 $h^{(2)}$ 的一致性可以由式(13)表示。

$$\min_{h^{(1)} \in H^{(1)}, h^{(2)} \in H^{(2)}} d_X(h^{(1)}, h^{(2)}) \quad (13)$$

式中, X 表示整个数据集, 包括标记和未标记数据; $h^{(1)}$ 和 $h^{(2)}$ 分别是假设空间 $H^{(1)}$ 和 $H^{(2)}$ 中的函数。在这种度量结构下, 通过最小化函数 $h^{(1)}$ 和 $h^{(2)}$ 的距离, 使函数 $h^{(1)}$ 和 $h^{(2)}$ 对数据的决策一致, 从而最小化 $h^{(1)}$ 和 $h^{(2)}$ 错误率的上界, 以寻找最优函数。本文用式(13)定义函数的一致性, 通过使用假设空间中的度量结构, 能够对不同的学习器使用不同的距离度量, 这样便能够针对具体问题给出更合适的一致性约束。

2.4 一种针对半监督协同训练的正则化算法

本文从正则化的角度研究多视图的半监督学习。通过对学习过程加以限制来控制假设空间的复杂度, 从而找到更满足要求的函数。具体地, 考虑两个层次的正则化, 即视图内的正则化和视图之间的正则化。在视图内的学习过程中, 通过正则化使学习得到的函数是其对应的假设空间中的最优函数; 在视图之间的协同学习过程中, 通过正则化使各个视图上的学习函数对未标记数据的预测相似。并且, 这两个层次的正则化同时进行。

对于视图内的正则化, 采用 2.2 节中提出的光滑性约束。在每个视图内的学习过程中惩罚学习函数的复杂性, 以得到在标记数据和未标记数据上具有相似预测性能的学习函数。通过辅助光滑函数 $\phi^{(i)}$, 使用假设空间的度量结构, 在假设空间 $H^{(i)}$ 中寻找与光滑函数 $\phi^{(i)}$ 在标记数据上的距离和在未标记数据上的距离相近的函数 $h^{(i)}$ 。

对于视图之间的正则化, 采用 2.3 节中提出的一致性约束, 同时惩罚各个视图上学习得到的函数的复杂性, 以使各个视图上学习得到的函数对数据的预测结果相近, 从而提高预测结果的可靠性。使用假设空间的度量结构, 在假设空间 $H^{(1)}, \dots, H^{(w)}$ 中分别寻找函数 $h^{(1)}, \dots, h^{(w)}$, 使 $h^{(1)}, \dots, h^{(w)}$ 之间的距离相近。通过一致性能进一步减少假设空间的大小。

综上所述, 本文提出一种多视图半监督学习中的正则化算法 (Multi-view Semi-supervised Regularization, 简称 MS-Regu 算法)。它在最小化经验误差的基础上, 同时考虑每个视图内学习函数的光滑性和多个视图上学习函数的一致性, 并采用统一的度量方法来衡量函数的这两种性质。用 $d_L(h^{(i)}, y)$ 表示函数 $h^{(i)}$ 与真实标记 y 的距离, 即经验误差。MS-Regu 算法的目标函数可以写为如下形式, 即

$$\sum_{i=1}^w (d_L(h^{(i)}, y) + (d_L(h^{(i)}, \phi^{(i)}) - d_U(h^{(i)}, \phi^{(i)}))^2) + \lambda \sum_{i,j=1}^w d_X(h^{(i)}, h^{(j)}) \quad (14)$$

通过最小化式(14), 得到每个视图上的最优函数 $h^{(1)}, \dots, h^{(w)}$ 。这些函数在尽可能一致的同时, 分别是假设空间 $H^{(1)}, \dots, H^{(w)}$ 中尽可能接近真实标记的函数。参数 λ 控制光滑性和一致性的作用比例, 当增大 λ 时, 求得的函数更加一致, 从而使算法的方差更小。MS-Regu 算法由两项构成, 第一项代表每个视图内的学习, 第二项代表视图之间的共同学习。其中, 每个视图内的学习包括经验误差和光滑性约束两方面的内容。在每个视图内学习的函数 $h^{(i)}$ 不仅应在标记数据上有小的经验误差, 而且应该具有光滑性。同时, 多个视图上的协同学习促使学习得到的函数 $h^{(1)}, \dots, h^{(w)}$ 具有一致性。

3 相关工作

对学习函数的光滑性进行约束是正则化算法中常用的方

法, 但不同的算法对光滑性有不同的定义, 例如传统的正则化方法^[6]的优化目标为

$$f^* = \operatorname{argmin}_{f \in H_K} \left\{ \frac{1}{L} \sum_{i=1}^L V(f(x_i), y_i) + \lambda \|f\|_K^2 \right\} \quad (15)$$

式中, 第一项为经验误差, 第二项为光滑性约束(也称为稳定性约束, 复杂性约束)。这样的光滑性约束没有考虑未标记数据的作用, 忽略了未标记数据中蕴含的信息。

半监督学习中的正则化方法通常在正则化项的设计中综合考虑未标记数据的作用。如在基于图的半监督学习^[10,11]中, 把未标记数据之间的近邻关系作为正则化项, 寻找在局部领域内标记具有平滑性的函数; 信息正则化方法^[12]把标记的复杂度作为正则化项, 寻找在数据分布稠密区域标记变化不大的函数。概括地说, 半监督学习中的正则化方法要求函数在标记和未标记数据上都具有光滑性。与传统正则化方法相比, 其对假设空间做了进一步的限制, 所以能找到更优的函数。

在多视图学习中, 一致性约束是减少假设空间复杂度的有效方法。一致性也可以称为相容性。标准协同训练算法中^[3], 对于两个视图上的半监督学习, 定义其相容性为: 若对于任意数据 $x = (x^{(1)}, x^{(2)})$, 有 $f^{(1)}(x^{(1)}) = f^{(2)}(x^{(2)})$, 则目标函数 $f = (f^{(1)}, f^{(2)})$ 与数据集 X 的分布 D 相容。Sindhwani 等人^[13]通过最小化两个视图上的学习器对未标记数据预测的不一致, 来提升学习算法的性能; Farquhar 等人^[14]提出监督 co-SVM 方法, 在最小化训练误差的同时最小化视图之间的不一致; Rosenberg 和 Bartlett^[15]研究一致性对模型误差和函数复杂性的影响, 并给出了它们之间的显式关系式。本文利用假设空间的度量结构统一度量学习函数的光滑性和一致性, 创新性地提出了一种针对半监督协同训练的正则化算法。

4 实验

WebKB 数据集是一个多视图的数据集, 是多视图半监督学习领域中常用的实验数据集, 最早由标准协同训练算法^[3]使用。它由一些网页数据组成, 这些网页来自于一些大学的网站。每个网页有两个视图, 一个视图由网页本身所包含的信息构成, 另一个视图由指向该网页的超级链接中所包含的信息构成。按照协同训练算法^[3]中的实验设置, 本文在 WebKB 数据集集中的 Course 数据集上进行实验。

Course 数据集由 1051 个网页构成, 其分为两类, 分别是课程网页和非课程网页, 其中课程网页占 22%。用 Rainbow 软件包^[16]对数据进行处理, 把每个网页用向量模型表示, 其中网页视图有 3000 个特征, 超级链接视图有 1840 个特征。对于这 1051 个网页组成的数据集, 将其随机划分为标记数据集 L 、未标记数据集 U 和测试集 TE , 其中 TE 占整个数据集的 25%, 剩下的 75% 是训练集 $X = LUU$ 。对于训练集 X , 将其按标记率 η 划分为 L 和 U , 即 η 代表标记数据在训练集中所占的比例。在实验中, 对 η 的不同取值进行了考察。注意在数据集的划分过程中保持类别比例不变。每个实验重复运行 10 次, 并取 10 次运行结果的平均值。

将本文中的 MS-Regu 算法与两个同类算法 vcb Regu 和 Co-training 在分类性能上进行对比。其中, Regu 代表传统的正则化方法^[6], Co-training 代表协同训练算法^[3]。实验结果如图 1 和图 2 所示。

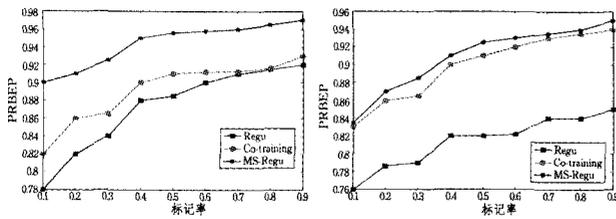


图 1 网页视图上的分类结果 图 2 超级链接视图上的分类结果

图 1 和图 2 分别是网页视图和超级链接视图上的分类结果。其中横坐标表示数据的标记率 η , 纵坐标表示准确率-召回率平衡点 (Precision-Recall Breakeven Point, 简称 PRBEP), 其计算方法为

$$\text{准确率} = \frac{\text{被判定为正例的数据中, 判定正确的数据个数}}{\text{被判定为正例的数据个数}}$$

$$\text{召回率} = \frac{\text{被判定为正例的数据中, 判定正确的数据个数}}{\text{正例数据总的个数}}$$

准确率也称为查准率, 召回率也称为查全率。准确率与召回率相等的点称为准确率-召回率平衡点。PRBEP 是两个类别的分类问题中常用的评价指标, PRBEP 值高意味着算法在分类准确的同时, 没有遗漏。在本文的实验中, 学习器的任务是将 Course 数据集分为两类, 即课程网页和非课程网页, 其中课程网页为正例, PRBEP 值越高, 算法性能越好。

从图 1 和图 2 可以看出, 标记率 η 从 0.1 增加到 0.9 的过程中, 3 个算法的性能都有所提高, 但 MS-Regu 算法始终比其它两个算法有更高的 PRBEP 值。分析其原因, 当标记率 η 增加时, 各个算法都能够从更多的标记数据中学习得到更好的决策函数, 但传统正则化算法在学习中不考虑未标记数据的信息, 协同训练算法不对决策函数的光滑性进行约束, 这可能导致忽略了数据中蕴含的一部分信息; 而 MS-Regu 算法综合考虑标记和未标记数据, 对决策函数的光滑性和一致性同时进行约束, 这样的学习机制一方面能够降低假设空间的复杂度, 另一方面能够使多个视图上的学习互相提供信息, 从而有助于学习到更好的决策函数。

从图 1 和图 2 中还可以看出, 网页视图上训练的分类器比超级链接视图上训练的分类器具有更高的 PRBEP 值, 平均高出 2 个百分点。分析其原因, 网页视图有 3000 个特征, 超级链接视图有 1840 个特征, 网页视图可能包含了更多有助于学习的信息。

结束语 从有限样本中学习往往是病态逆问题, 其解具有不稳定性。在半监督学习中标记数据的数量较少, 仅从标记数据中学习往往不能得到适定的解。经典机器学习理论认为, 解决此问题的方法是通过正则化对学习过程加以限制, 控制假设空间的复杂度, 将不适定问题转化为适定问题, 以使学习器具有更好的泛化性能。

本文对多视图半监督学习中的正则化问题进行了研究, 创新性地提出了一种两个层次的正则化算法 (MS-Regu 算法)。MS-Regu 算法针对多视图的半监督学习, 综合利用未标记数据提供的信息和数据多个视图提供的信息定义函数的光滑性和一致性。MS-Regu 算法在每个视图内的学习过程中, 用光滑性作为正则化项, 限制每个视图内学习的函数具有光滑性; 在多个视图的协同学习过程中, 用一致性作为正则化项, 限制多个视图中学习到的函数对同一数据的预测相似。

通过假设空间中的度量结构度量函数的光滑性和一致性, 并在学习过程中同时考虑视图内学习器的光滑性和视图之间学习器的一致性。

本文在 WebKB 真实数据上进行了实验。结果显示, MS-Regu 算法比仅考虑光滑性或仅考虑一致性的正则化算法具有更好的分类性能。

参考文献

- [1] Olivier C. Introduction to semi-supervised learning[M]//Chappelle O, Scholkopf B, Zien A, eds. Semi-Supervised Learning. Cambridge: MIT Press, 2006
- [2] 周志华, 尹学松, 肖宇, 等. 半监督学习专刊[J]. 软件学报, 2008, 19(11): 2789-2868
- [3] Avrim B, Tom M. Combining labeled and unlabeled data with co-training[C]//Proceedings of the 11th Annual Conference on Learning Theory. Madison, WI, 1998: 92-100
- [4] 周志华. 半监督学习中的协同训练风范[M]//周志华, 王珏. 机器学习及其应用. 北京: 清华大学出版社, 2007: 259-275
- [5] Zhou Zhi-hua, Li Ming. Semi-supervised learning by disagreement[J]. Knowledge and Information Systems, 2010, 24 (3): 415-439
- [6] Vapnik V N. Statistical Learning Theory[M]. New York: Wiley-Interscience, 1998
- [7] Dale S, Finnegan S. Metric-based methods for adaptive model selection and regularization[J]. Machine Learning, 2002, 48(1-3): 51-84
- [8] de SV. Learning classification with unlabeled data[M]. Advances in Neural Information Processing Systems 6, 1994: 112-119
- [9] Sanjoy D, Littman M L, David M. PAC generalization bounds for co-training[M]. Advances in Neural Information Processing Systems 14, Cambridge, MA: MIT Press, 2002: 375-382
- [10] Mikhail B, Partha N, Vikas S. On manifold regularization[C]//Proceedings of the 10th International Workshop on Artificial Intelligence and Statistics. Savannah Hotel, Barbados, 2005: 17-24
- [11] Zhang Tong, Alexandrin P, Byron D. Linear prediction models with graph regularization for web-page categorization[C]//Proceedings of the 12th ACM SIGKDD international conference on knowledge discovery and data mining. Philadelphia, PA, USA ACM Press, 2006: 821-826
- [12] Corduneanu A, Jaakkola T. On information regularization[C]//Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence. 2003
- [13] Vikas S, Partha N, Mikhail B. A co-regularized approach to semi-supervised learning with multiple views[C]//Proceedings of Working Notes of the ICM'05 Workshop on Learning with Multiple Views. Bonn, Germany, 2005
- [14] Farquhar J D R, Hardoon D, Meng H, et al. Two view learning: SVM-2K, theory and practice[M]. Advances in Neural Information Processing Systems 18, 2006
- [15] David R, Bartlett P L. The rademacher complexity of co-regularized kernel classes[C]//Proceedings of the 11th International Conference on Artificial Intelligence and Statistics. 2007
- [16] Andrew M. Bow; A toolkit for statistical language modeling, text retrieval, classification and clustering[OL]. <http://www.cs.cmu.edu/~mccallum/bow>, 1996