基于 Tri-training 的半监督 SVM

李昆仑,张 伟,代运娜
LI Kun-lun,ZHANG Wei,DAI Yun-na

河北大学 电子信息工程学院,河北 保定 071002
College of Electronic and Information Engineering,Hebei University,Baoding,Hebei 071002,China
E-mail:wzhanghbu@gmail.com

LI Kun-lun,ZHANG Wei,DAI Yun-na.Semi-supervised SVM based on Tri-training.Computer Engineering and Applications,2009,45(22):103-106.

Abstract: One of the main difficulties in machine learning is how to solve large-scale problem effectively,and the labeled data are limited and fairly expensive to obtain.In this paper a new semi-supervised SVM is proposed.It applies Tri-training to improve SVM.The semi-supervised SVM uses a few labeled data to train few initial SVM classifiers and makes use of the large number unlabeled data to modify the classifier iteratively.Experiments on UCI dataset show that Tri-training can improve the classification accuracy of SVM and can increase the difference of classifier,the accuracy of final classifier will be higher.Although Tri-training doesn't put any constraints on the supervised learning algorithm,the proposed method uses the SVMs with three different kernel functions as the supervised learning algorithm.The different kernel can increase the difference of the three SVMs,so the performance of co-training will be better.Theoretical analysis and experiments show that the proposed algorithm has excellent accuracy and speed of classification.

Key words: semi-supervised learning;co-training;Tri-training;Support Vector Machine(SVM);least square support vector machine

摘 要:当前机器学习面临的主要问题之一是如何有效地处理海量数据,而标记训练数据是十分有限且不易获得的。提出了一种新的半监督 SVM 算法,该算法在对 SVM 训练中,只要求少量的标记数据,并能利用大量的未标记数据对分类器反复的修正。在实验中发现,Tri-training 的应用确实能够提高 SVM 算法的分类精度,并且通过增大分类器间的差异性能够获得更好的分类效果,所以 Tri-training 对分类器的要求十分宽松,通过 SVM 的不同核函数来体现分类器之间的差异性,进一步改善了协同训练的性能。理论分析与实验表明,该算法具有较好的学习效果。

关键词:半监督学习;协同训练;Tri-training;支持向量机;最小二乘支持向量机
DOI:10.3778/j.issn.1002-8331.2009.22.034 **文章编号**:1002-8331(2009)22-0103-04 **文献标识码**:A **中图分类号**:TP181

1 引言

随着数据采集与存储技术的飞速发展,研究者拥有大量的数据,但已标记数据却相当有限,大量的未标记数据被搁置起来。如垃圾邮件处理,面对海量电子邮件,得到足够数量的已标记邮件是不可能实现的,所以大部分的邮件还是要作为未标记数据,不能帮助算法提高学习性能。其他如语音识别、网页处理等领域,都存在同样的问题。

传统的监督学习需要一组足够多的已标记数据作为训练集,否则无法获得足够泛化性能的监督学习方法,而在实际应用中,得到大量标记数据是非常困难的,甚至无法实现;而无监督学习,试图通过发现未标记数据中的隐含结构,从而构造出相应的学习器,这导致无监督学习通常很难保证较高的学习精度[1-2]。在这种情况下,利用传统机器学习策略不能得到足够泛化性能和精度的学习器[1]。

半监督学习作为一种近年新提出的学习策略,弥补了监督学习与无监督学习的不足, 同时利用了标记数据和未标记数据。半监督学习已经成为该领域的研究热点,吸引着越来越多的学者对其进行深入地研究,该理论及其实现算法也因此得到了快速的发展[1-3]。

提出基于协同训练改进策略——Tri-training 的半监督 SVM 算法,充分利用了未标记数据,并通过实验验证了该算法具有较好的学习性能。

2 支持向量机和最小二乘支持向量机

2.1 支持向量机

支持向量机(SVM)是一种建立在统计学习理论基础上的

基金项目:国家自然科学基金(the National Natural Science Foundation of China under Grant No.60773062);河北省自然科学基金项目(the Natural Science Foundation of Hebei Province of China under Grant No.F2009000215);河北省科技支撑计划项目(the Science and Technology Supporting Program of Hebei Province under Grant No.072135188);河北省教育厅科研计划项目(the Scientific Research Project of Department of Hebei Education of China under Grant No.2008312)。

作者简介:李昆仑(1962-),男,博士,副教授,硕士生导师,主要研究领域为模式识别、人工智能等;张伟(1982-),男,硕士生,主要研究领域为模式识别。

收稿日期:2008-06-24 **修回日期**:2008-09-16

通用机器学习算法。与现有的其它机器学习算法相比,具有以下特点[4]:(1)良好的泛化性能;(2)能够得到全局最优解;(3)核技巧的应用;(4)很好的鲁棒性[4-5]。

以二类分类问题为例简单介绍 SVM[6-7],假设有一组训练数据集$\{\boldsymbol{x}_i,y_i\}(i=1,2,\cdots,N)$,其中 $\boldsymbol{x}_i\in R^n$ 是第 i 个训练样本,y_i 是 $\boldsymbol{x}_i$ 相应的类标号+1 或-1。SVM 通过解决最优化问题:

$$\begin{cases}\min\ \phi(\boldsymbol{w})=\dfrac{1}{2}\|\boldsymbol{w}\|^2+C\sum\limits_{i=1}^{N}\xi_i\\ \text{s.t. } y_i[(\boldsymbol{w}^{\mathrm{T}}\cdot\varphi(\boldsymbol{x}_i))+b]-1+\xi_i\geqslant 0,i=1,2,\cdots,N\\ \xi_i\geqslant 0\end{cases}\tag{1}$$

获得最优分类超平面。

2.2 最小二乘支持向量机

最小二乘支持向量机(LS-SVM)是 Suykens 等在 1999 年提出的经典 SVM 的一种改进算法。其基本思想是通过一组线性等式代替 SVM 中的二次规划问题,从而大大提高了程序运行速度[6]。

LS-SVM 是经典 SVM 基础上的改进,所以它的最优化问题与经典 SVM 是相似的,但约束条件变为等式形式:

$$\begin{cases}\min\ \phi(\boldsymbol{w})=\dfrac{1}{2}\|\boldsymbol{w}\|^2+\dfrac{1}{2}\gamma\sum\limits_{i=1}^{N}e_i^2\\ \text{s.t. } y_i[(\boldsymbol{w}^{\mathrm{T}}\cdot\varphi(\boldsymbol{x}_i))+b]-1+e_i=0,i=1,2,\cdots,N\end{cases}\tag{2}$$

3 半监督学习及协同训练策略

3.1 半监督学习

半监督学习试图在已标记的训练样本有限的情况下,利用大量的未标记样本,在不增加对未标记数据处理耗费的同时提高机器学习算法的学习精度。半监督学习思想最早可追溯到二十世纪五六十年代[2],而后诞生了称为自学习(Self-learning)或自训练(Self-training)的半监督分类方法。随着研究的不断深入,目前已有很多新的方法,如生成模型(Generative Models)、自训练(Self-training)、协同训练(Co-training)以及基于图的方法(Graph-Based Method)[2-3]等,半监督学习的应用也扩展到了回归及聚类等领域。

3.2 协同训练策略 Co-training 和 Tri-training

Co-training 是 Blum 和 Mitchell 于 1998 年提出的一种半监督算法。简单地说,协同训练策略就是假设数据有两个不同的充分冗余视图(view),而每个视图的属性集合都足以单独训练出一个强分类器。在这两个视图的基础上分别利用少量的已标记数据训练两个分类器,再用训练得到的两个分类器分别对未标记数据进行预测,并从中挑选出置信度较高的帮助对方重新训练分类器,以改善性能[1,3,8]。

由于现有的大部分数据都不能满足充分冗余视图的条件,所以 Goldman 和 Zhou 提出了一种 Co-training 的改进算法[9],此算法不再需要充分冗余视图,取而代之的是利用两个不同类型的分类器完成学习,但要求在每轮迭代中采用 10 重交叉验证,以确定未标记样本的标记置信度,因此该方法十分耗时。周志华等在 2005 年针对 Co-training 及其改进算法存在的问题提出了 Tri-training 算法[10]。Tri-training 同样没有要求充分冗余视图,而在分类器的设置上采用了三个分类器进行协同训练,这样既利用了多分类器的协同优势又避免了传统协同策略验证时间长、对分类算法及样本类型要求苛刻的不足。邓超等于 2007 年提出了基于自适应数据剪辑策略的 Tri-training 算法[11]。

4 半监督 SVM

Co-training 对数据属性的苛刻要求及交叉验证带来的巨大时间损耗,都令其难于推广到实际应用中。因此选择应用 Tri-training 策略,同时对三个 SVM 分类器进行训练。虽然 Tri-training 相对于 Co-training 增加了一个分类器,但根据集成策略可知,更多的独立分类器,会使最终分类器的集成效果更好[12],同时第三个分类器的引入还有效的避免了 Co-training 中的交叉验证过程,从而大幅提高了整体算法的运行速度。Tri-training 在应用于 SVM 及 LS-SVM 算法后,既提高了 SVM 算法对未标记数据的学习效果又体现出了 LS-SVM 算法速度快的特点。

4.1 基于噪声数据学习的基本理论

假设 $\sigma=\{(x_1,y_1),(x_2,y_2),\cdots,(x_m,y_m)\}$表示抽取出的包含 m 个数据的训练样本序列,其中 $\boldsymbol{x}_i$ 均来自于整体数据集,y_i 为数据的标记。L_i 表示可能的假设,令 $F(L_i,\sigma)$表示 L_i 与序列 σ 不一致性样本的个数。则有以下结论[13]。

定理 若抽取的序列 σ 的规模 m 满足:

$$m\geqslant\frac{2}{\varepsilon^2(1-2\eta)^2}\ln\left(\frac{2N}{\delta}\right)\tag{3}$$

且 L_i 是令 $F(L_i,\sigma)$最小的假设,则有:

$$\Pr[d(L_i,L_*)\geqslant\varepsilon]\leqslant\delta\tag{4}$$

这表明带噪声训练集假设具有 PAC 可学习性。

其中 ε 是误差容忍度,也就是最坏假设情况下的分类误差率,δ 是置信度参数,η 是训练集噪声率上界,范围为 $0\leqslant\eta\leqslant 0.5$,$N$ 是假设数目,$d(L_i,L_*)$是假设 L_i 和真实假设 L_* 间均衡差别(symmetric difference)的概率之和。

令 $C=2\mu\ln(2N/\delta)$其中 μ 是使式(3)等号成立的值,则

$$m=\frac{C}{\varepsilon^2(1-2\eta)^2}$$

$$u=\frac{C}{\varepsilon^2}=m(1-2\eta)^2\tag{5}$$

4.2 基于 Tri-training 的半监督 SVM 算法

Tri-training 需要三个分类器来进行协同训练,但对这三个分类器并没有特别的要求。由集成策略知,分类器间的差异性越大则最后的集成效果也越好,同时由于在实际应用中,大多数情况下数据的属性是未知的,不知道何种核函数更适合于相应数据,所以在这里为适应不同类型数据同时也能够增大三个分类器间的独立性,所选择的三个经典 SVM 分类器或 LS-SVM 分类器均采用了不同的核函数。这也最大程度的避免半监督算法退化为三个自学习(Self-training)分类器的集成,而且虽然三个 SVM 分类器应有了不同的核函数,但输出形式是相同的,这也易于最后的集成决策。

假设初始的标记样本集为 L,样本数为$|L|$,未标记样本集为 U,大小为$|U|$。三个 SVM 分类器首先分别利用样本集 L 的 Bootstrap 重采样后数据进行训练。训练完成后,在后面每轮再训练过程中,三个分类器中的一个作为训练对象,另外两个作为辅助分类器,两个辅助分类器对 U 中样本进行分类,并将相同标记意见的样本与相应的标记组成集合 L',再利用$|L\cup L'|$重新训练第一个 SVM 分类器。需要注意的是,L'在下一轮的优化过程中并不作为标记数据处理,而是被重新放回 U 中,作为未标记数据在下一轮中重新使用。若 L'中的样本被辅助分类器预测正确,则对被训练分类器是增加了一个正确的训练样本,

但也有可能会产生预测错误的样本,其对于被训练分类器就是增加了一个噪声。噪声必然会对被训练分类器带来一定的不利影响,如何尽量消除噪声带来的影响呢?由上一节的定理可知,假设 L''为新一轮中用于训练第一个分类器的样本集,η_L 为 L 的分类噪声率,e'和 η'表示上一轮中两辅助分类器的分类误差率上界和被优化分类器的分类噪声率,可知有:

$$\eta'=\frac{\eta_L|L|+e'|L'|}{|L\cup L'|} \tag{6}$$

再由式(5)知,若 $u''>u'$即

$$|L\cup L''|\left(1-2\frac{\eta_L|L|+e''|L''|}{|L\cup L''|}\right)>|L\cup L'|\left(1-2\frac{\eta_L|L|+e'|L'|}{|L\cup L'|}\right) \tag{7}$$

则 $\varepsilon''<\varepsilon'$。由此可知随着未标记样本的加入,分类器性能会提高,并且加入的未标记样本越多,性能提高的也越多,由此可知虽有噪声加入,但数据量的增加能够抵消噪声所带来的影响,这也是为何要将每次从未标记数据集中提取出的数据,在下一轮中仍作为未标记数据的原因[10-11,13]。

选择经典 SVM 与 LS-SVM 两种分类器,是为了观察半监督策略应用在分类精度较好的经典 SVM 算法和在分类速度上较有优势的 LS-SVM 算法上的性能提升效果,并根据实际应用领域的不同给出相应的选择建议。

5　实验结果及分析

实验中,使用了 4 组来自于 UCI[14]的数据集:Australian、German、Ionosphere、Wdbc,验证算法的有效性,它们都是二类分类数据,数据量分别是 690、1 000、351 和 569,其中 Australian 数据集为一组 14 维的数据,但其中有 5%的缺失信息;German 是一组 24 维的完整数据集;Ionosphere 是一组 34 维的完整数据集;Wdbc 为一组 32 维完整线性可分数据集。

在实验中,SVM 算法采用的是 OSU SVM3.0 工具箱和 LS-SVM 的 Matlab 工具箱。SVM 和 LS-SVM 的核函数分别选择:线性核 $K(\boldsymbol{x},\boldsymbol{x}_i)=(\boldsymbol{x}\cdot\boldsymbol{x}_i)$,多项式核 $K(\boldsymbol{x},\boldsymbol{x}_i)=(\boldsymbol{x}\cdot\boldsymbol{x}_i+c)^d$ 和 RBF 核 $K(\boldsymbol{x},\boldsymbol{x}_i)=\exp\left(-\frac{\|\boldsymbol{x}-\boldsymbol{x}_i\|^2}{\sigma^2}\right)$,并且经典 SVM 与 LS-SVM 采用相同的协同半监督策略。

经典 SVM 中 RBF 核参数:σ^2=1;多项式核参数:d=3;惩罚系数 C 均设为 10。

LS-SVM 中 RBF 核参数:σ^2=1;多项式核参数:d=3;规范化系数(regularization parameter)gam 均设为 1。

实验环境:P4 2.4 GHz CPU,256 MB 内存;WINDOWS XP 系统,MATLAB V7.1 版。

5.1　实验结果

实验中数据做以下处理:数据被随机分为三部分,每部分都与整体具有相似的分布状态即正负样本比例。全部数据的 25%作为测试数据集,剩余数据作为训练数据,并在实验中按不同比例分为两部分,一部分数据作为已标记数据集 L,另一部分作为未标记数据集 U。

采用不同数量的未标记数据进行实验,是为观察标记数据与未标记数据在不同比例下,半监督算法对 SVM 分类器性能提升的程度。为了体现算法实验效果的一般性,实验会在不同比例下分割三次,并将算法重复运行 3 次,获得 3 次结果,而最终结果取 3 次结果的平均值。

对于半监督 LS-SVM 算法选择初始误差率、最终误差率和程序运行时间作为算法性能的评价指标(其中初始的误差率指仅利用 L 的 Bootstrap 采样后数据训练得到的三个初始分类器,对测试数据集的集成分类结果),对于半监督经典 SVM 算法,则选择最终误差率和程序运行时间作为实验结果。

实验结果如表 1~表 5,表中结果精确到小数点后 2 位,其中 SVM-L 表示三个 SVM 分类器皆采用线性核函数时,SVM-D 表示三个 SVM 分类器采用三个不同核函数时。若程序运行时间超过 14 小时,则不记录最后结果。

表 1　80%的未标记率

数据集	LS-SVM		SVM-L*	SVM-D*
	初始误差率/(%)	最终误差率/(%)	最终误差率/(%)	最终误差率/(%)
Australian	15.12	14.92	13.950	–
German	28.67	27.20	26.800	26.800
Ionosphere	21.96	18.94	14.394	10.984
Wdbc	15.96	11.74	7.400	9.620

表 2　60%的未标记率

数据集	LS-SVM		SVM-L	SVM-D
	初始误差率/(%)	最终误差率/(%)	最终误差率/(%)	最终误差率/(%)
Australian	14.54	13.76	113.37	–
German	28.13	26.27	25.60	24.80
Ionosphere	21.21	16.67	12.50	8.33
Wdbc	11.03	8.69	5.40	9.39

表 3　40%的未标记率

数据集	LS-SVM		SVM-L	SVM-D
	初始误差率/(%)	最终误差率/(%)	最终误差率/(%)	最终误差率/(%)
Australian	13.18	12.98	12.79	–
German	26.13	25.73	25.20	25.20
Ionosphere	17.80	10.23	10.61	6.82
Wdbc	9.39	6.81	5.16	7.04

表 4　20%的未标记率

数据集	LS-SVM		SVM-L	SVM-D
	初始误差率/(%)	最终误差率/(%)	最终误差率/(%)	最终误差率/(%)
Australian	12.43	12.40	12.21	–
German	26.93	25.07	24.40	–
Ionosphere	15.15	10.23	9.50	5.30
Wdbc	7.04	6.10	4.93	–

表 5　程序运行时间

未标记率/(%)	Australian/s			German/s			Wdbc/s		
	LS-SVM	SVM-L	SVM-D	LS-SVM	SVM-L	SVM-D	LS-SVM	SVM-L	SVM-D
80	62.68	496.83	–	128.64	46.44	46.88	95.92	5.48	53.05
60	65.49	10 040.10	–	142.70	174.88	1 187.34	97.64	111.82	178.88
40	68.35	18 476.63	–	145.93	293.33	10 463.89	98.71	232.62	596.13
20	72.71	24 933.93	–	161.15	463.34	–	105.04	329.44	–

5.2 实验结果分析

由表 1~表 3 可知,在分类误差率方面,应用半监督策略后不同未标记比例下,SVM 算法的分类精度都有较大幅度的提高,且采用三个不同核函数的半监督经典 SVM 方法优于采用线性核函数时,而采用线性核函数的半监督经典 SVM 方法又优于采用三个不同核函数的 LS-SVM 半监督方法。这说明提出的基于 Tri-training 半监督 SVM,能够通过采用不同核函数以提高分类器之间的差异性,从而提高算法在最后集成后的分类精度,而且最后的集成输出也十分方便。经过半监督训练后的经典 SVM 分类效果虽优于半监督 LS-SVM,但二者的差距并不大。

表 5 给出了算法的时间损耗,三者顺序正好与依据精度排序的结果相反。半监督 LS-SVM 算法的速度明显快于半监督 SVM 算法,而且数据量越大优势也越明显。随着未标记率的降低,参与三个分类器初始化训练的样本量逐渐增多,所有算法的运行时间都会越来越长。对于一般实验数据,最快与最慢算法间的差距约 3 倍(20%未标记率下)。而对于特征缺失数据最大差距达到了约 343 倍(Australian 数据集),对此类数据 SVM 算法需反复调整分类面的方向与偏置即 $\boldsymbol{w}$ 和 b 的值,以获得最佳分类超平面,因而导致了算法训练时间变长。但综合实验数据可以看出,半监督 LS-SVM 的误差率并没有随速度优势的增大而大幅的增高。

6 总结

提出了一种基于 Tri-training 半监督 SVM 方法。在分类精度上,半监督 SVM 方法较半监督 LS-SVM 有一定的优势,所以基于 Tri-training 半监督经典 SVM 适合于对实时性要求不高的应用领域,如文本分类、语音识别等,能够获得较好的学习精度;而基于 Tri-training 半监督 LS-SVM 在精度上虽略有不足,但能够保证较高的运行速度,且在大数据集上的表现尤为明显,所以更适合对实时性要求较高的应用领域,如实时入侵检测。

半监督学习策略的引入,能够有效提高算法的学习精度,说明半监督 SVM 确实具有研究与实用价值。但由于 Tri-training 策略中存在大量反复训练过程,会耗费一定时间,所以在进一步的工作中,如何避免重复性的训练,是一个亟待解决的问题,增量学习方式可以作为一种可能解决方案引入到该算法中,以解决重复计算问题。

参考文献:

[1] 周志华,王珏.机器学习及其应用[M].北京:清华大学出版社,2007.
[2] Chapelle O,Schölkopf B,Zien A.Semi-supervised learning[M].[S.l.]:The MIT Press,2006.
[3] Zhu X J.Semi-supervised learning literature survey,Technical Report 1530[R].Department of Computer Sciences,University of Wisconsin at Madison,Madison,WI,2007-12.
[4] Vapnik V.The nature of statistical learning theory [M].New York:Springer-Verlag,2000.
[5] Steve R G.Support vector machines classification and regression[R].Department of Electronics and Computer Science,University of Southampton,1998.
[6] Suykens J A,Vandewalle J.Least squares support vector machine classifiers[J].Neural Processing Letters,1999,9(3):293-300.
[7] Cristianini N,Shawe-Taylor J.An introduction to support vector machines and other Kernel-based learning methods[M].[S.l.]:Cambridge University Press,2000.
[8] Blum A,Mitchell T.Combining labeled and unlabeled data with co-training[C]//Proceedings of the 11th Annual Conference on Computational Learning Theory,Madison,WI,1998:92-100.
[9] Goldman S,Zhou Y.Enhancing supervised learning with unlabeled data[C]//Proceedings of the 17th ICML.San Francisco,CA:Morgan Kaufmann,2000:327-334.
[10] Zhou Z H,Li M.Tri-training:Exploiting unlabeled data using three classifiers[J].IEEE Transactions on Knowledge and Data Engineering,2005,17(11):1529-1541.
[11] 邓超,郭茂祖.基于自适应数据剪辑策略的 Tri-training 算法[J].计算机学报,2007,30(8):1214-1226.
[12] Duda R O,Hart P E.Pattern classification [M].2nd ed.New York:Wiley,2001.
[13] Angluin D,Laird P.Learning from noisy examples [J].Machine Learning,1988,2(4):343-370.
[14] UCI repository of machine learning databases[EB/OL].http://www.ics.uci.edu/mlearn/MLRepository.html.

(上接 65 页)

进行时钟的同步,它测量的只是单个源发出的包之间的相对延迟量,而 Coates 的网络合并要求进行测量时需要成对探测源在发送探测时进行一定误差范围内的时钟同步工作;最后,基于链路延迟属性的网络拓扑合并算法是根据链路延迟属性值进行合并节点的推断,算法更加直观并且容易实现。

4 结语

讨论了在一个大规模通讯网络中,如何对多个逻辑网络拓扑进行合并的问题。研究的成果在于改进了三明治探测方案和 SCT 算法,并在此基础上提出了基于链路延迟属性的网络拓扑合并的新方案。此方案主要优点有:测量对网络带来的数据流量小;探测方法不需要时钟同步;基于链路延迟属性的网络拓扑合并算法要更加直观,可行性好。

从实验结果来看,方案解决了单源测量在网络逻辑拓扑判定的研究中,所获得的逻辑拓扑只能推断出源端点到接收端点之间树形的拓扑结构,而不能正确反映网状拓扑的问题。实验中将不同源端点测量所得的逻辑拓扑作合并处理,所得到结果接近于实际的网络拓扑。但在大规模网络测量中,如何减少测量结果受到网络噪音及流量的干扰,以及如何减少测量数据本身对网络流量的影响等问题是在以后工作中需要进一步研究的。

参考文献:

[1] Castro R,Coates M,Liang G,et al.Network tomography:Recent developments[J].Statistical Science,2004,19(3):499-517.
[2] Coates M,Rabbat M,Nowak R.Merging logical topologies using end-to-end measurements[C]//ACM Internet Measurement Conference,Miami,FL,October 2003.
[3] Coates M,Castro R,Nowak R,et al.Maximum likelihood network topology identification from edge-based unicast measurements[C]//Proc ACM SIGMETRICS.New York:ACM Press,2002:11-20.
[4] 张巍,王郁武.系统聚类树算法在网络拓扑判定中的研究[J].四川大学学报:自然科学版,2008,45(6):1332-1336.
[5] Casto R M,Coates M J,Nowak R D.Likelihood based hierarchical clustering[J].IEEE Trans on Signal Process,2004,52(8):2308-2321.