

基于最小最大模块化集成特征选择的改进

周 丰,王未央

(上海海事大学 信息工程学院,上海 201306)

摘要:随着数据规模的扩大,单个弱分类器的准确率已经无法很好地对未知样本进行预测,为此提出了集成学习。在集成学习与分类器结合的同时,集成的思想同样被用到了特征选择中。从提高对样本预测的准确率的角度出发,提出一种基于最小最大模块化(Min-Max-Module, M3)的策略。它同时将集成学习应用到了特征选择算法和分类器中,并对比了四种集成策略以及三种不同的分类方法。结果表明,提出的方法在大多情况下能取得不错的效果,并且能很好地处理不平衡的数据集。

关键词:特征选择;集成学习;最小最大模块化策略;不平衡数据

中图分类号:TP391

文献标识码:A

文章编号:1673-629X(2016)09-0149-05

doi:10.3969/j.issn.1673-629X.2016.09.033

Improvement of Multi-classification Integrated Selection Based on Min-Max-Module

ZHOU Feng, WANG Wei-yang

(School of Information & Engineering, Shanghai Maritime University, Shanghai 201306, China)

Abstract: With the expansion of the data size, a single weak classifier has been unable to predict unknown samples accurately. To solve this problem, an integrated learning is proposed. Combined the integrated learning and classification, the idea of integration is also used in the feature selection at the same time. For the increase of sample prediction accuracy, a strategy based on Min-Max-Module (M3) is put forward. It makes integrated learning applied to feature selection algorithms and classifier, and compares four kinds of integration strategies as well as three different classification methods. The results show that the proposed method can be able to achieve good results in most cases, and can well handle imbalanced data sets.

Key words: feature selection; integrated learning; Min-Max-Module (M3); Imbalance Data Sets (IDS)

1 概 述

随着各个领域涌现出的大量数据,机器学习与数据挖掘已经被应用到各行各业。但是现实生活中的数据非常复杂,平滑干净的数据非常难得,大部分的数据中都有缺失值、异常值等噪声。此时,数据预处理显得格外重要。为了提高机器对这些数据的处理能力,可以通过一定的数学方法(如牛顿插值法、拉格朗日插值法等)进行处理,而特征选择作为降低数据维度、平滑噪声的一种有效方法,成为了研究热点^[1]。

在数据挖掘中,海量的原始数据中存在着大量不完整、不一致、有异常的数据,它们会严重影响到数据挖掘建模的执行效率,所以数据清洗尤为重要。数据

清洗主要是删除原始数据集中的无关数据、重复数据、平滑噪声数据,筛选掉与挖掘主题无关的数据,处理缺失值、异常值等^[2]。处理缺失值和异常值,通常有删除记录、数据插补和不处理等方法,而异常值一旦被检测到,往往可以当作缺失值处理。常用的方法有均值/众数/中位数插补法、固定值法、最近邻插补、回归方法以及插值法。插值法一般有拉格朗日插值法和牛顿插值法,文中使用牛顿插值法。箱型图分析提供了识别异常值的标准:异常值通常被定义为小于 $Q_L - 1.5IRQ$ 或大于 $Q_U + 1.5IRQ$ 的值。其中, Q_L 为下四分位数; Q_U 为上四分位数; IRQ 为四分位数间距, $IRQ = Q_U - Q_L$ 。

收稿日期:2015-09-17

修回日期:2016-01-06

网络出版时间:2016-08-23

基金项目:国家自然科学基金青年项目(61303100)

作者简介:周 丰(1991-),女,硕士研究生,研究方向为挖掘算法中的特征处理;王未央,硕士生导师,研究方向为数据库系统、系统与数据整合、数据挖掘及其在港航、海洋、物流信息系统中的应用。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20160823.1343.032.html>

特征选择是从特征集合中挑选出满足一定评价准则的特征子集的过程,特征选择过程可以除去不相关的冗余特征^[1-2],从而达到降维的目的。特征选择算法一般分为四类:过滤器、封装器、嵌入式和组合。过滤器最大的特点是直接从特征固有的性质出发来评判特征的重要性,并不考虑分类器,该类方法具有较高的效率,如 Fisher^[3]、Relief^[4]等。封装器则是依赖于分类器,它采用分类器来评价性能,这种方法采用搜索策略来寻找最优特征子集,其搜索策略有前向搜索、后向搜索、随机搜索等,准确率较高但是效率较低。嵌入式则是在构建分类器的过程中进行特征选择。而组合式先采用过滤器去除一些特征,再对剩下的特征子集采用封装器进行搜索,结合了过滤器和封装器的优点。

目前,数据量在不断增长,传统的特征选择方法的效率显然已经无法跟上数据增长的脚步^[1]。为了更好地处理大规模数据,Hoi^[5]等结合在线特征选择和一种嵌入式算法,将原始数据转换为序列型的数据;Wu^[6]等使用在线流式特征选择的方法。以上两种方法都能明显地提升对大规模数据的处理能力。

文中使用的是集成的方法,主要步骤包括:结合箱型图分析和牛顿插值法对数据进行缺失值和异常值分析;再通过某种策略将数据划分,从而将原始特征选择和分类任务划分为多个较小的相互独立的可并行计算的子任务,每一个子任务同时进行特征选择以及分类算法;最后利用最小最大策略将分类结果进行集成。

2 集成学习

2.1 集成学习的概念与框架

与传统的单个学习模型相比,集成学习则是通过同时构建多个不同的基学习模型,并使用某种策略把多个模型的学习结果进行组合,从而获得最终的学习结果。

此外,集成思想与特征选择相结合,用以提高特征选择的稳定性^[2,7-9]。大量的理论研究和实际应用表明,集成学习有利于构建性能更好的学习模型。集成学习在分类问题中大致分为两个步骤。首先,根据数据集训练出多个不同的分类器;然后,将未知数据在不同分类器上的预测结果通过某种策略进行汇总,整合成最终的预测结果。汇总的策略有投票法、权重法、聚类法等。

常见的集成策略有以下几种,如 Bagging、Ada-Boost、M3、RandomSpace 等。该实验使用基于 M3 的集成策略对结果进行集成,特征选择算法使用了 Fisher 与 ReliefF;分类算法使用经典的、准确率较高的支持向量机,并与单个支持向量机分类器和朴素贝叶斯三者之间进行比较。

2.2 集成的策略

在每个子任务结束后,需要将每个任务的结果以某种方法集成起来,文中将介绍与对比四个集成的策略:均值法、投票法、K-中心聚类法以及最小最大集成策略。此处主要介绍前三种:

投票法对于特征选择而言是先将结果转换为特征子集,然后统计每个特征被选中的情况,将出现次数最多的 M 个特征作为最终输出特征;同理对于分类而言则是统计每一个子分类器对同一个样本预测的结果,取票数最高的作为输出;

均值法将子任务返回的结果进行线性相加取均值得到最后的输出;

K-中心聚类集成是应用了聚类的思想,从多个子任务中选择具有代表性的结果作为输出。相比前两种方法,该方法可以保护特征之间的关联性。文中采用了 1-中心聚类集成。

3 最小最大集成策略

基于最小最大模块化 (Min-Max-Module, M3) 的分类集成策略最早由 Lu^[10]等提出。该策略主要包含两个步骤:任务分解和分类结果的合成。

3.1 任务分解

在任务分解阶段,对于一个 K 类的分解问题,首先采用“一对一”的策略将其分解为 $K(K-1)/2$ 个二分类问题。假设 K 类的训练数据集表示为:

$$X = \{(X_i^l, y^i)\}_{i=1}^{L_i}, i = 1, 2, \dots, K$$

其中, L_i 表示第 i 类样本的个数; x_i^l 表示第 i 类样本中的第 l 个样本; y^i 表示第 i 类样本对应的标签。

那么通过“一对一”的策略,第 i 类样本和第 j 类样本组成的二分类问题的训练数据集可以表示为:

$$X_{i,j} = \{(x_i^l, y^i)\}_{i=1}^{L_i} \cup \{(x_j^l, y^j)\}_{j=1}^{L_j}, i, j = 1, 2, \dots, K, i \neq j$$

如果二分类问题的规模较大或者具有不平衡性,可以进一步将它们划分成规模更小的较为平衡的子问题。

任务分解方法包括基于随机的分解方法和基于超平面的分解方法。该实验中使用基于超平面的方法,其具体过程如算法 1 所示。假设把 C_i 类的训练样本分解成 N_i 个子集,把 C_j 类的训练样本分解成 N_j 个子集。这样就可以把 C_i 类和 C_j 类的二分类问题分解成 $N_i \times N_j$ 个子二分类问题进行解决^[11]。

算法 1: 基于超平面的数据集分块方法。

输入: 某 k 类问题第 i 类的训练样本 $X_i, i = 1, 2, \dots, k$ 。

(1) 计算 C_i 类的每个训练样本 x 与超平面 $Z_1 + Z_2$

+ ⋯ + Z_n = 0 的距离。

$$\text{dist}(x, H) = \frac{1 \times x_1 + 1 \times x_2 + \cdots + 1 \times x_n}{\sqrt{1^2 + 1^2 + \cdots + 1^2}} = \frac{x_1 + x_2 + \cdots + x_n}{\sqrt{n}}$$

其中, $x_j, j = 1, 2, \cdots, n$ 是样本 x 的分量。

(2) 根据已经计算的 $\text{dist}(x, H)$ 的值, 对 C_i 类的训练样本进行排序, 即把 C_i 类的训练样本按空间分布进行排序。

(3) 把已经排序的训练样本按前后顺序划分成 N_i 份, 每一份的训练样本个数近似相等 (相差不超过一个), 即把该类训练样本的分布空间分割成 N_i 个部分, 但不是分割后的空间相等, 而是保证各个部分空间中包含的样本个数相等。

3.2 分类结果集成

通过对第 i 类样本和第 j 类样本组成的二分类问题进行进一步的划分, 从而得到 $N_i \times N_j$ 个二分类子问题, 然后在每个子问题对应的训练数据集上训练相应的分类器, 得到 $N_i \times N_j$ 个基分类器, 表示为:

$$C_{i,j}^{u,v} = \text{Train}(X_{i,j}^{u,v}), u = 1, 2, \cdots, N_i, v = 1, 2, \cdots, N_j$$

对于测试样本, 使用这些基分类器分别对其进行预测, 得到预测标签:

$$O_{i,j}^{u,v} = \text{Test}(C_{i,j}^{u,v}, e)$$

对于预测结果, 分别采用最小规则和最大规则进行合成:

MIN 规则: 是对拥有相同正类训练样本集和不同负类训练样本集的分类结果取最小值;

MAX 规则: 是对拥有相同负类训练样本集和不同正类训练样本集的分类结果取最大值。

该分类集成策略的整体流程如下:

算法 2: 基于最小最大规则的集成算法。

输入: 训练集 X , 测试样本 e , 第 i 类样本的划分块数 N_i , 数据划分方法 P ;

输出: 测试样本的预测标签 O 。

训练与测试阶段:

将 K 类样本 X 划分为 X_1, X_2, \cdots, X_k

for $i = 1, 2, \cdots, K - 1$

for $j = i + 1, \cdots, K$

{ $x_i^1, x_i^2, \cdots, x_i^{N_i}$ } = $P(X_i)$

{ $x_j^1, x_j^2, \cdots, x_j^{N_j}$ } = $P(X_j)$

end for

end for

for $u = 1, 2, \cdots, N_i$

for $v = 1, 2, \cdots, N_j$

$X_{i,j}^{u,v} = \text{划分数据}_j^{(u,v)}$

$$C_{i,j}^{u,v} = \text{Train}(X_{i,j}^{u,v})$$

$$O_{i,j}^{u,v} = \text{Test}(X_{i,j}^{u,v}, e)$$

$$O_{i,j} = \text{Max}(O_{i,j}^1, O_{i,j}^2, \cdots, O_{i,j}^{N_j})$$

end for

end for

$$O = \text{Voting}(O_{i,j}), i = 1, 2, \cdots, K - 1, j = i + 1, \cdots, K$$

3.3 最小最大策略的集成特征选择以及分类

传统的基于最小最大策略的分类是先将数据集进行特征选择等预处理, 再利用上文提到的数据划分方法将数据划分为 $M \times N$ 个样本子集, 再将数据块进行合并分类, 对每一个样本都有 $M \times N$ 个预测标签, 再利用最大最小策略可以得到每一个样本最终的预测标签作为输出^[12]。

传统的基于最小最大策略的集成特征选择则是先使用上文提到的数据划分方法进行数据划分, 得到 $M \times N$ 个样本子集, 然后在每个样本子集上进行特征选择, 得到 $M \times N$ 个特征选择结果。最后利用最小化集成单元和最大化集成单元对这多个特征选择结果进行组合^[12-15]。

文中提出的方法是结合以上两种传统的集成方法, 将特征选择和分类同时使用最小最大模块化进行集成, 先进行数据分块, 对每一个数据子集进行特征选择, 并优先对特征选择集成。同时, 保存划分的数据子集, 在得到最优特征子集之后, 更新数据子集并对新的数据子集进行分类集成, 从而得到最后的结果。其过程如图 1 所示。

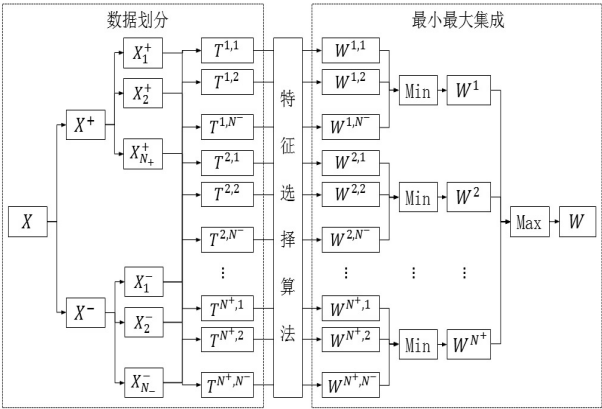


图 1 基于最小最大规则集成分类特征选择的框架

4 实验及仿真

4.1 实验步骤

文中所做的相关工作及实验步骤为:

(1) 对原始数据做箱型图处理, 寻找异常值并删除, 再将排除完异常值的数据集运用牛顿插值法 (完成异常值和缺失值的处理过程具体见算法 1)。

(2) 接下来对数据进行标准化、归一化处理, 区间

为[0,1]。

(3) 对处理好的数据进行交叉验证或者直接进行训练集和测试集的按比例划分。

(4) 对训练样本和测试样本分别使用基于超平面的方法得到若干较小数据子集,对每个数据子集使用相同的特征选择算法,得到降维之后的数据。

(5) 用降维之后的数据更新之前的数据子集,并对每个数据子集使用支持向量机算法进行分类。

(6) 根据最小最大规则对分类结果进行集成得到最终输出。

本节通过实验结合了基于最小最大策略的分类算法和特征选择算法,分别对比 ReliefF 和 Fisher 特征选择算法、投票法、均值法、K-中心聚类法以及最小最大策略对分类准确率的影响。在集成方面,将文中提出的方法与传统的 M3-SVM 和 M3-Naïve Bayes 方法进行对比。实验数据集为 PCMAC 和 Adult。

4.2 实验准备

(1) 实验数据集。

PCMAC 数据集包含 1 943 条样本,每条样本有 3 290 维属性,包含了若干异常值和缺失值,标签一共有两类,是一个低数据量高维的样本。文中对其进行交叉验证,得到训练样本和测试样本。该数据集具体的分布情况如表 1 所示。

表 1 PCMAC 数据集分布情况

类型标签	样本个数	所占比例/%	M3 中分块个数
-1	961	49.46	5
+1	982	50.54	5

Adult 数据集包含 32 561 条样本,124 维属性,同时包含了异常值和缺失值,一共有两类标签。文中训练集包含 22 696 条样本,测试集 9 865 条样本,具体分布如表 2 和表 3 所示。

表 2 Adult 训练集数据分布情况

类型标签	样本个数	所占比例/%	M3 分块个数
-1	17 190	75.74	6
1	5 506	24.26	2

表 3 Adult 测试集数据分布情况

类型标签	样本个数	所占比例/%	M3 分块个数
-1	7 530	76.33	6
1	2 335	23.67	2

需要说明的是,由于 M3 希望尽可能地保持每个子数据块的样本数相似,故对 M3 的分块个数需要针对样本个数区分,这样也能很好地处理不平衡数据集。

(2) 分类器的选择及评价准则。

文中采用数据分类器算法是支持向量机。支持向量

机有良好的学习能力和泛化能力,主要思想是:对于样本的输入空间,构造一个最优的超平面,使得超平面到两类样本之间的距离最大化。它遵循结构风险最小化的原则,使得错误概率上界最小化,因此还可以有效减小过拟合。在支持向量机中,文中采用高斯核函数,其 Sigmod 值设置为 2,损失函数 C 设置为 32,采用 SMO 算法计算其参数。

对于分类结果的评价采用错误率度量标准,但是错误率不考虑类别之间的不平衡。对于平衡数据,常采用准确率或错误率来衡量;对于不平衡数据,常采用的评价标准包括 ROC 曲线、AUC、F-Measure 和 G-Mean 等。文中采用 G-Mean。

错误率 = $\frac{FN + TP}{TP + FN + FP + TN}$

PR = $\frac{TP}{TP + FN}$

NR = $\frac{TN}{TN + FP}$

准确率(Precision) = $\frac{TP}{TP + FP}$

G-Mean = $\sqrt{PR \times NR}$

其中,TP 表示正类样本被正确分类的个数;FN 表示正类样本被错误分类的个数;TN 表示负类样本被正确分类的个数;FP 表示负类样本被错误分类的个数。

4.3 实验结果

4.3.1 在 PCMAC 数据集上的实验结果

该数据集样本数较少,维数较高,且为平衡数据集,在该样本上使用了 Fisher 特征选择算法。

(1) 使用 ReliefF 特征选择算法。

基于超平面划分数据后,对每一个数据子集使用 Fisher 特征选择算法后使用 SVM 作为分类器,分别采用四种不同的集成策略以 G-Mean 为标准进行对比,如图 2 所示。

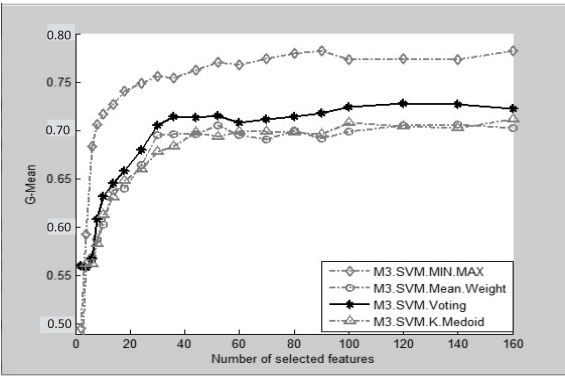


图 2 不同集成策略对比图(1)

对文中提出的集成特征选择(FSE)结合集成分类(CFE)与传统的 FSE 进行对比,并通过传统的 FSE 对比了 SVM 和 NB 两个分类器的效果,如图 3 所示。

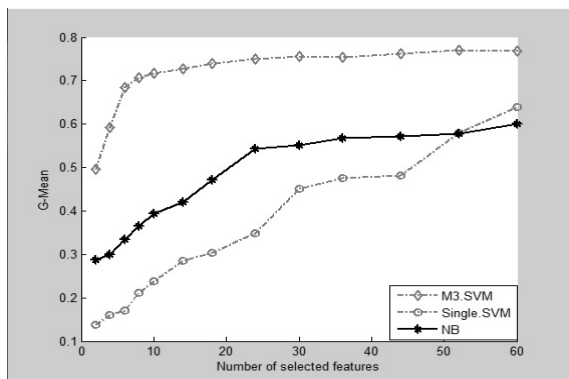


图3 集成算法与单个算法对比图(1)

4.3.2 在 Adult 数据集上的实验结果

该数据集样本数较多,且不平衡,属性个数较少,此处仅采用了 ReliefF 特征选择算法,同样对比了文中提出的 FSE+CFE 与传统的集成特征选择,实验结果如图 4 和图 5 所示。

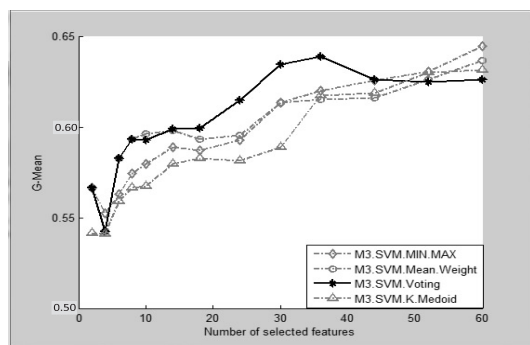


图4 不同集成策略对比图(2)

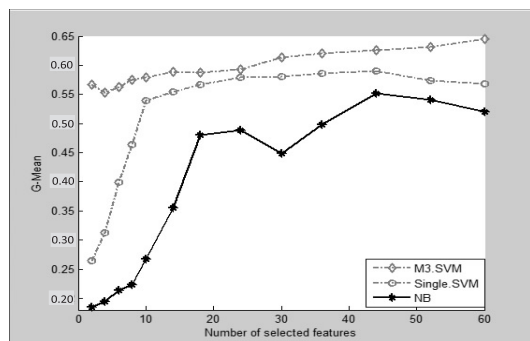


图5 集成算法与单个算法对比图(2)

4.3.3 实验结果分析

实验主要对比了四种集成策略:均值法、投票法、K-中心聚类法以及基于最小最大策略(分别记为 Mean. Weight、Voting、K. Medoid、MIN. MAX)。并对比了三种不同的分类方法:基于最小最大策略的集成特征选择与集成分类结合方法、传统的集成特征选择结合 SVM、传统的集成特征选择结合朴素贝叶斯(分别记为 M3. SVM、SVM、NB)。数据划分部分使用了超平面划分,特征选择算法使用了 ReliefF 和 Fisher,使用 G-Mean 作为评价准则。实验数据集使用 PCMAC 和 Adult,前者样本较少、特征较多,故采用 10 次交叉验证。

根据实验结果,在样本较少特征较多的 PCMAC 中,文中提出的 M3-FS 结合 M3-SVM 的效果明显优于其他集成策略和分类方法,特别是选取 60 至 80 维特征。对于大量低维的样本 Adult,由图 4、图 5 可见,四种集成策略具有相似的结果及趋势;在分类器选择方面,文中提出的 M3-FS 结合 M3-SVM 的效果优于 M3-FS 结合 SVM 以及 M3-FS 结合 NB。同时可以看出,使用 M3 的正负类样本分别分块策略也能很好地处理 Adult 这个不平衡的数据集。

综上所述,从分类准确率的角度,文中提出的方法在大部分情况下均优于其他方法。

5 结束语

文中提出了一种对传统的集成特征选择方法加以改进的方法,即将分类集成与特征选择的集成结合起来,以有效地提高对大规模数据的处理能力。该方法通过基于超平面的方法将数据划分成多个数据子集,将原来的任务转换为可同时进行的多个独立的子任务,然后使用最小最大集成单元对分类结果进行集成,得到最终的预测标签。通过比较,实验结果表明在四种集成策略中,最小最大集成策略是占有一定的优势的。在面对传统的特征选择集成方法中,在准确率方面,结合分类集成的方法的效果更为理想。

参考文献:

- [1] Tang J L, Alelyani S, Liu H. Feature selection for classification; a review [M]. Florida: The Chemical Rubber Company Press, 2013.
- [2] Li Y, Gao S, Chen S. Ensemble feature weighting based on local learning and diversity [C]//Proc of AAAI conference on artificial intelligence. [s. l.]: [s. n.], 2012.
- [3] Gu Q, Li Z, Han J. Generalized fisher score for feature selection [C]//Proceedings of the twenty-seventh conference on uncertainty in artificial intelligence. Barcelona, Spain: [s. n.], 2011.
- [4] Robnik-Šikonja M, Kononenko I. Theoretical and empirical analysis of ReliefF and RReliefF [J]. Machine Learning, 2003, 53(1-2): 23-69.
- [5] Hoi S C H, Wang J, Zhao P, et al. Online feature selection for mining big data [C]//Proc of international workshop on big data, streams and heterogeneous source mining: algorithms, systems, programming models and applications. [s. l.]: ACM, 2012: 93-100.
- [6] Wu X, Yu K, Wang H, et al. Online streaming feature selection [C]//Proc of international conference on machine learning. [s. l.]: [s. n.], 2010: 1159-1166.
- [7] Woznica A, Nguyen P, Kalousis A. Model mining for robust

= 8 时,TPE 和 RZF 的性能已经非常接近,而此时天线数 $M = 128$,用户数 $K = 32$ 。

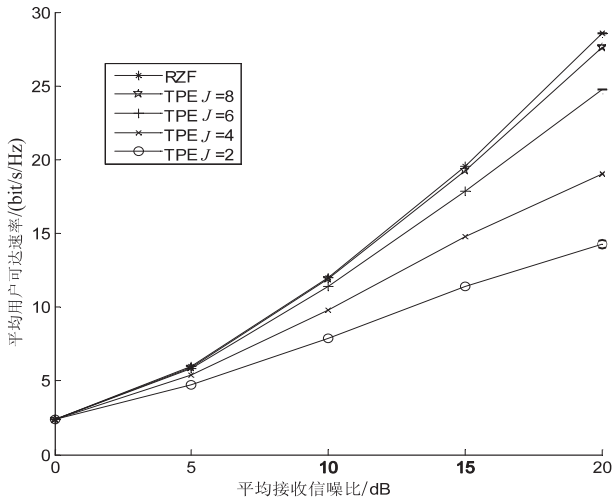


图2 不同阶数用户可达速率与信噪比的关系

5 结束语

文中从系统吞吐量的角度研究了单小区多用户下行链路中降低线性预编码复杂度的方法。采用基于截断多项式扩展预编码方法建立了一种系统吞吐量和预编码复杂度的平衡。通过截断多项式扩展等效算法实现了多级硬件并行流水结构,并采用这种并行处理算法消除了系统的额外延迟,同时可以随着信噪比的高低动态增加或减少多项式阶数 J 的值。文中推导出了可以使吞吐量最大化的最优预编码系数,通过仿真验证了所提算法在较低复杂度下性能良好。

参考文献:

[1] Marzetta T L. Noncooperative cellular wireless with unlimited numbers of base station antennas[J]. IEEE Transactions on Wireless Communications,2010,9(11):3590-3600.

[2] 李兴旺.无线通信系统中的大规模MIMO关键理论及技术

研究[D].北京:北京邮电大学,2015.

[3] 戚晨皓,黄永明,金石.大规模MIMO系统研究进展[J].数据采集与处理,2015,30(3):544-551.

[4] Hachem W, Khorunzhy O, Loubaton P, et al. A new approach for capacity analysis of large dimensional multi-antenna channels[J]. IEEE Transactions on Information Theory, 2008, 54(9):3987-4004.

[5] 吴伟陵,牛凯.移动通信原理[M].第2版.北京:电子工业出版社,2009.

[6] 张贤达.矩阵分析与应用[M].北京:清华大学出版社,2005.

[7] Sessler G, Jondral F. Low complexity polynomial expansion multiuser detector for CDMA systems[J]. IEEE Transactions on Vehicular Technology, 2005, 54(4):1379-1391.

[8] 徐国珍,蒋伟,刘安,等.有限反馈下大规模MIMO系统性能仿真[J].重庆邮电大学学报:自然科学版,2012,24(5):589-594.

[9] Hoydis J, Debbah M, Kobayashi M. Asymptotic moments for interference mitigation in correlated fading channels[C]//Proc of ISIT. [s. l.]:[s. n.], 2011:2796-2800.

[10] 邱玲.多用户多小区MIMO通信技术[M].北京:人民邮电出版社,2011.

[11] Joham M, Utschick W, Nossek J A. Linear transmit processing in MIMO communications systems[J]. IEEE Transactions on Signal Processing, 2005, 53(8):2700-2712.

[12] Bjornson E, Zakhour R, Gesbert D, et al. Cooperative multi-cell precoding: rate region characterization and distributed strategies with instantaneous and statistical CSI[J]. IEEE Transactions on Signal Processing, 2010, 58(8):4298-4310.

[13] 杨大成.移动传播环境理论基础[M].北京:机械工业出版社,2003.

[14] Bjornson E, Jorswieck E. Optimal resource allocation in coordinated multi-cell systems[J]. Foundations and Trends in Communications and Information Theory, 2013, 9(2-3):113-381.

(上接第153页)

feature selection[C]//Proc of ACM SIGKDD conference on knowledge discovery and data mining. [s. l.]: ACM, 2012: 913-921.

[8] Awada W, Khoshgoftaar T M, Dittman D, et al. A review of the stability of feature selection techniques for bioinformatics data [C]//Proc of international conference on information reuse and integration. [s. l.]:[s. n.], 2012:356-363.

[9] 季薇,李云.基于局部能量的集成特征选择[J].南京大学学报:自然科学版,2012,48(4):499-503.

[10] Lu B L, Ito M. Task decomposition and module combination based on class relations; a modular neural network for pattern

classification[J]. IEEE Transactions on Neural Networks, 1999, 10(5):1244-1256.

[11] 周国静,李云.基于最小最大策略的集成特征选择[J].南京大学学报:自然科学版,2014,50(4):457-465.

[12] 陈晓明.海量高维数据下分布式特征选择算法的研究与应用[J].科技通报,2013,29(8):79-81.

[13] 连惠城.最小最大模块化网络及人脸属性分类研究[D].上海:上海交通大学,2008.

[14] 解男男.机器学习方法在入侵检测中的应用研究[D].长春:吉林大学,2015.

[15] 闫国虹.支持向量机不平衡问题和增量问题算法研究[D].西安:西安电子科技大学,2012.