

摘要: 为了快速、准确地判别田间籽棉的成熟度, 提取了描述棉瓣形状的15个结构特征, 基于10折交叉验证比较了封装器下穷举搜索并基于封装器停止搜索(WE-W)和过滤器下启发式搜索并基于封装器停止搜索(FH-W)这二种特征选择算法的执行效率和分类性能。分别以验证集上Bayes分类器的误分率(WE-W)和训练集上的类可分性测量值(FH-W)为评价函数, 在训练集上穷举搜索(WE-W)和启发式搜索(FH-W)最优特征子集,  $l=1, 2, \dots, 15$ , 并于Bayes分类器在验证集上的平均误分率极小时停止搜索(WE-W和FH-W)。结果显示, WE-W和FH-W算法在预测集上 $l=3$ 处分别获得了85.39%(WE-W)和85.28%(FH-W)的平均识别率, 表明FH-W算法执行效率高、分类性能好, 对实际应用有参考意义。

关键词: 籽棉成熟度 封装器 穷举搜索 过滤器 启发式搜索 特征选择

### Comparison of two Feature Selection Algorithms Oriented to Raw Cotton Ripeness Discrimination

WANG Ling, LIU De-ying, JI Chang-ying\*

College of Engineering, Nanjing Agricultural University/Jiangsu Province Engineering Lab for Modern Facility Agriculture Technology & Equipment, Nanjing 210031, China

Abstract: To discriminate the ripeness of cotton quickly and accurately, 15 shape structure features were extracted from cotton images and the execute efficiencies and classification accuracy of their feature subset selection algorithms such as Wrapper-based Exhaustive searching and Wrapper-based stopping(WE-W) and Filter-based Heuristic searching and Wrapper-based stopping(FH-W) were compared by using 10-fold cross-validation. By taking the error rate of a Bayes classifier on validation set (WE-W) and the class-separability measuring value on a training set (FH-W) as assessing functions, the optimal  $l$  ( $l=1, 2, 3, \dots, 15$ ) feature subset was searched by using exhaustive (WE-W) and heuristic (FH-W) strategies on the training set, which stops at the minimum error rate of Bayes-classifier on the validation set(WE-W and FH-W). Experimental results show that the average classification rates of WE-W and FH-W algorithms on the prediction set are 85.39% (WE-W) and 85.28% (FH-W) at  $l=3$ , respectively. It concludes that the FH-W algorithm can be a reference in practice for its higher execute efficiency and good classification accuracy.

Keywords: Cotton ripeness Wrapper Exhaustive search Filter Heuristic search Feature selection

收稿日期 2013-02-06 修回日期 2013-04-06 网络版发布日期 2013-08-20

基金项目:

国家“十五”863计划资助项目;江苏省农机 基金资助

通讯作者: 王玲

作者简介: 王玲(1966-), 女, 江西南昌人, 博士, 副教授, 1988年于南京航空航天大学获得学士学位, 2006年、2009年于南京农业大学分别获得硕士、博士学位, 主要从事图像处理与模式识别技术方面的研究。

作者Email: Lingw@njau.edu.cn

#### 参考文献:

- [1]孙伟艳. 模式分类中特征选择问题的研究[D]. 哈尔滨: 哈尔滨理工大学, 2009. SUN W Y. Research on feature selection for pattern classification [D]. Harbin: Harbin University of Science and Technology, 2009. (in Chinese)
- [2]NG W W Y, YEUNG D S, FIRTH M, et al.. Feature selection using localized generalization error for supervised classification problems using RBFNN [J]. Pattern Recognition, 2008, 41(12): 3706-3719.
- [3]RUIZ R, RIQUELME J C, AGUILAR-RUIZ J S. Incremental wrapper-based gene selection from microarray data for cancer classification [J]. Pattern Recognition, 2006, 39(12): 2383-2392.
- [4]LIANG J N, YANG S, WINSTANLEY A. Invariant optimal feature selection: A distance discriminant and feature ranking based solution [J]. Pattern Recognition, 2008(5), 41: 1429-1439.
- [5]LIU Y, ZHENG Y F. FS\_SFS: A novel feature selection method for support vector machines [J]. Pattern Recognition, 2006, 39(7): 1333-1345.
- [6]SEBBAN M, NOCK R. A hybrid filter/wrapper approach of feature selection using information theory [J]. Pattern Recognition, 2002, 35(4): 835-846.
- [7]DONG M, KOTHARI R. Feature subset selection using a new definition of classifiability [J]. Pattern Recognition Letters, 2003, 24(9-10): 1215-1225.
- [8]HUA J P, TEMBE W D, DOUGHERTY E R. Performance of feature-selection methods in the classification of high-dimension data [J]. Pattern Recognition, 2009, 42(3): 409-424.
- [9]NAKARIYAKUL S, CASSASANT D P. An improvement on floating search algorithms for feature subset selection [J]. Pattern Recognition, 2009, 42(9): 1932-1940.
- [10]李先锋, 朱伟兴, 纪滨, 等. 基于图像处理和蚁群优化的形状特征选择与杂草识别[J]. 农业工程学报, 2010, 26(10): 178-182. LI X F, ZHU W X, JI B, et al.. Shape feature selection and weed recognition based on image processing and ant colony optimization [J]. Transactions of the CSAE, 2010, 26(10): 178-182. (in Chinese)
- [11]董岩, 张涛, 李文明, 等. 机载立体测绘相机滚转轴伺服系统的辨识与设计[J]. 光学 精密工程, 2011, 19(7): 1580-1587. DONG Y, ZHANG T, LI W M, et al.. Identification and design of roll axis servo system in airborne stereo mapping camera[J]. Opt. Precision Eng., 2011, 19(7): 1580-1587. (in Chinese)
- [12]董超, 田联房. 最速上升关联向量机高光谱影像分类[J]. 光学 精密工程, 2012, 20(6): 1398-1405. DONG CH, TIAN L F. Hyperspectral image classification by steepest ascent relevance vector machine[J]. Opt. Precision Eng., 2012, 20(6): 1398-1405. (in Chinese)
- [13]WEBB A R. 统计模式识别 [M]. 第二版. 王萍 等译. 北京: 电子工业

出版社, 2004. WEBB A R. Statistical Pattern Recognition [M]. 2nd ed. Wang Ping et al. translation. Beijing: Publishing House of Electronics Industry, 2004. (in Chinese) [14]FRIEDMAN N, GEIGER D, GOLDSZMIDT M. Bayesian network classifiers [J]. Machine Learning, 1997, 29(2-3): 131-163. [15]宋枫溪, 高秀梅, 刘树海, 等. 统计模式识别中的维数削减与低损降维[J]. 计算机学报, 2005, 28(11): 1915-1922. SONG F X, GAO X M, LIU SH H, et al.. Dimensionality reduction in statistical pattern recognition and low loss dimensionality reduction [J]. Chinese Journal of Computers, 2005, 28(11): 1915-1922.(in Chinese)

本刊中的类似文章

1. 李伟红, 龚卫国, 杨利平. 人脸特征选择中的SVM泛化误差估计[J]. 光学精密工程, 2008, 16(8): 1452-1458

---

Copyright by 光学精密工程