Learning Hybrid Kernel Machine Ensemble for An Imbalanced Data Problem and Colonoscopic Image Analysis

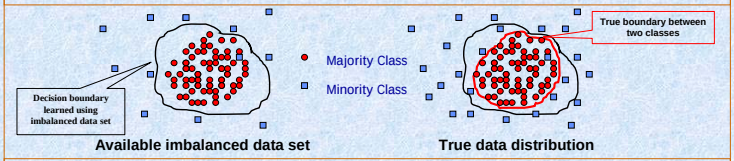
Peng Li
P.Li@cs.ucl.ac.uk
Department of Computer Science, University College London, Gower St., London, WC1E 6BT, UK.
This is author's previous work as a Ph.D. candidate in Nanyang Technological University, Singapore.

Summary

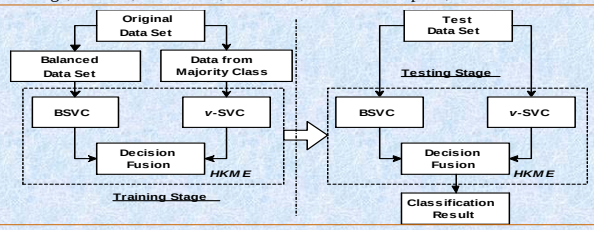
A two-class imbalanced data problem (IDP) emerges when the data from majority class are compactly clustered and the data from minority class are scattered. Though a discriminative binary Support Vector Machine (SVM) can be trained by manually balancing the data, its performance is usually poor due to the inadequate representation of the minority class. A recognition-based one-class SVM can be trained using the data from the well-represented class only. However, it is not highly discriminative. Exploiting the complementary natures of the two types of SVMs in an ensemble can bring benefits from both worlds in addressing the IDP. The proposed Hybrid Kernel Machine Ensemble (HKME) was later used for detecting abnormal regions in colonoscopic images.

HKME for Imbalanced Data Problem

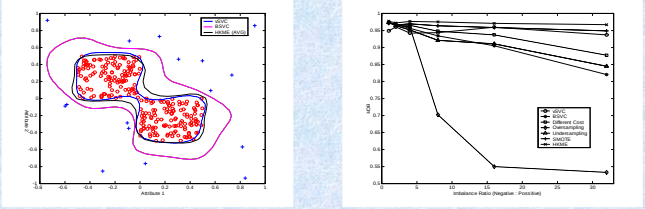
- **Imbalanced Data Problem (IDP)**:
  - **Problem**: Decision boundary is pushed biased to the minority class
  - **Reason:** Inadequate representation of the minority class



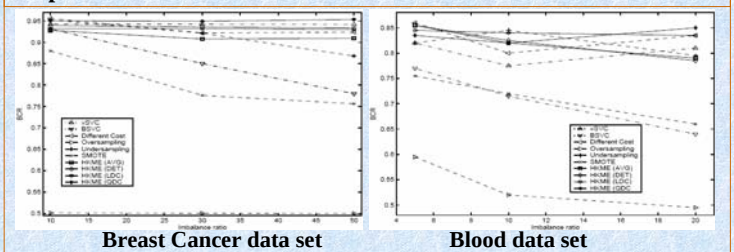
- **Two-class classifier -** Binary SVM is a discriminative classifier
  - **Good** in discrimination **but** Suffers from imbalanced data problem
- **One-class classifier -** One-class SVM is a recognition-based classifier.
  - **Good** for address imbalanced data problem **but** not very discriminative leaving one class data unused
- **Exploiting the complementary natures of these two kernel machines using an ensemble (HKME) is expected to improve the classification over either of them separately.**
- **Ensemble rules:**
  - Average, Product, Maximum, Minimum, Decision Template, Stacked Generalization etc.



•Experimental results on checkerboard toy data sets



•Experimental results on two UCI data sets



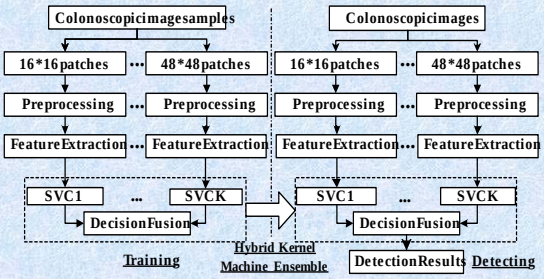
Breast Cancer data set | Blood data set

HKME for Colonoscopic Image Analysis

- **Difficulties in abnormal region detection**
  - Only a short interval of the colonoscopic image sequences shows abnormalities - **Imbalanced data problem – HKME**
  - Abnormal regions often do not occupy the whole images and show different location, shape , color and size
    - **Image region representation problem**
- **Solution: Image region representation based on multi-size patches**
  - Larger size patches:
    - Better classification accuracy rate
    - Small abnormal regions may be ignored.
  - Smaller size patches:
    - Good for representing the abnormal regions
    - Poorer classification due to less information contained.

Multiple-size patches are used simultaneously, At least some among all patch size can better characterize the image regions

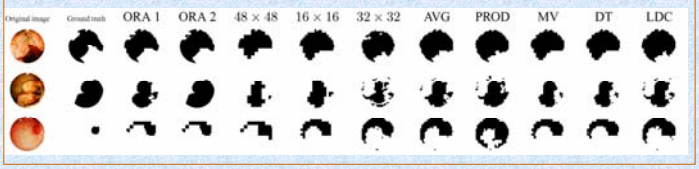
Proposed Framework



•Experimental results

•Data Set: 58 color colonoscopic images with the resolution of 256 × 256 pixels. Pixels in the original images were manually labeled to provide the ground truths.

Fusion rules	S1	S2	S3	AA	AB	A+1	A+2	A+3	A+12	A+13	A+23	ALL
AVG	0.551	0.556	0.563	0.751	**0.551**	**0.761**	**0.761**	**0.754**	**0.769**	**0.765**	**0.762**	0.565
PROD	0.551	0.556	0.563	0.745	0.535	**0.768**	0.659	0.730	0.598	0.656	0.572	0.548
MV	-	-	-	0.751	**0.551**	0.751	0.751	**0.751**	**0.765**	**0.765**	**0.765**	0.704
DT	**0.744**	**0.738**	**0.745**	0.753	0.538	0.753	0.753	**0.753**	0.753	0.753	0.753	0.751
LDC	**0.746**	**0.741**	**0.745**	**0.763**	0.533	**0.765**	**0.764**	**0.756**	**0.765**	**0.763**	**0.763**	**0.764**
Oracle	0.903	0.900	0.899	0.868	0.569	0.953	0.953	0.953	0.953	0.953	0.953	0.955



Conclusion

A new hybrid kernel machine ensemble was proposed for a type of imbalanced data sets. Benefited from a discriminative binary SVM and a recognition-based one-class SVM, the proposed hybrid kernel machine ensemble performs better than using either of them separately in the imbalanced data sets. It was then used to detect abnormal regions in colonoscopic images. Experimental results show the feasibility of the proposed method. The proposed method can also be extended to other applications such as object detection, face detection, etc.