CLASSIFICATION - ERROR COST MINIMIZATION STRATEGY: dCMS

Devi Parikh, Tsuhan Chen

Department of Electrical and Computer Engineering, Carnegie Mellon University

Motivation

- **Several applications have different costs associated with different classification-errors**
  Example: intrusion detection, biometric recognition, etc.
- **Most classification systems are geared towards minimizing the error rate and not cost**
  True objective function to be minimized is the cost of classification-error and not error-rate itself
- **Existing approaches can not handle multi-class problems or dynamically changing costs**
  ROC curves (multi-class? [1]) ; cost-sensitive Adaboost [2] (dynamically changing costs?)

Goal

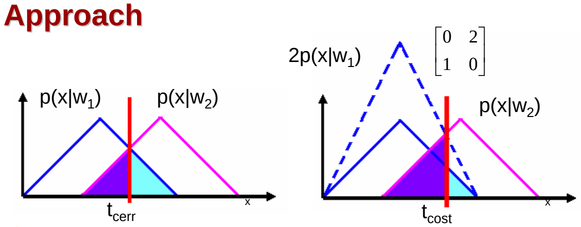
Develop a classification-error cost minimization strategy that

- **Can deal with multiple classes in a principled manner**
- **Is a simple post-training step**
  Does not require re-training of classifiers for changing costs
- **Is classifier type independent**
  Exploits statistical properties of the trained classifier

Contributions

- **Statistically significant reduction in costs incurred**
- **Effective on**
  - **a variety of applications**
  - **data sets of varying dimensionalities**
  - **a variety of classifier types**

Approach



- **Solution for a two-class, one-feature problem, known distributions**
- **If unknown distribution**
  Estimate with a histogram
- **If multiple-features**
  Classification system: maps multiple-features to a single score/feature
- **If multiple-classes**
  High dimensional histogram is not feasible … so then?

Intuition: Convert C-class problem to C 2-class problems

We have a trained classification system

Given Cost matrix:

$$\begin{bmatrix} 0 & cst_{12} & \dots & cst_{1C} \\ cst_{21} & 0 & \dots & cst_{2C} \\ \dots & & & \\ cst_{C1} & cst_{C2} & \dots & 0 \end{bmatrix}$$

Current Confusion matrix:

$$\begin{bmatrix} cnf_{11} & cnf_{12} & \dots & cnf_{1C} \\ cnf_{21} & cnf_{22} & \dots & cnf_{2C} \\ \dots & & & \\ cnf_{C1} & cnf_{C2} & \dots & cnf_{CC} \end{bmatrix}$$

Probability of a misclassified instance classified as class c actually belonging to class i:

$$p_{ic} = \frac{cnf_{ic}}{\sum_{j=1}^{C} cnf_{jc},\ j \neq c},\ i \neq c$$

Expected cost of false positives:

$$E(FPcst_c) = \sum_{i=1}^{C} p_{ic} \text{ x } cst_{ic}$$

$$\begin{bmatrix} cnf_{11} & cnf_{12} & \dots & cnf_{1C} \\ cnf_{21} & cnf_{22} & \dots & cnf_{2C} \\ \dots & & & \\ cnf_{C1} & cnf_{C2} & \dots & cnf_{CC} \end{bmatrix}$$

Current Confusion matrix

$$\begin{bmatrix} 0 & cst_{12} & \dots & cst_{1C} \\ cst_{21} & 0 & \dots & cst_{2C} \\ \dots & & & \\ cst_{C1} & cst_{C2} & \dots & 0 \end{bmatrix}$$

Cost matrix

$$\begin{bmatrix} 0 & Exp(FNcst_1) \\ Exp(FPcst_1) & 0 \end{bmatrix} \quad \begin{bmatrix} 0 & Exp(FNcst_2) \\ Exp(FPcst_2) & 0 \end{bmatrix} \quad \vdots \quad \begin{bmatrix} 0 & Exp(FNcst_C) \\ Exp(FPcst_C) & 0 \end{bmatrix}$$

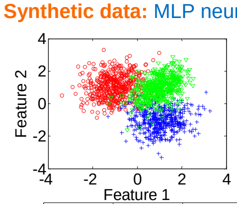
(Iterate to get a new confusion matrix with new thresholds)

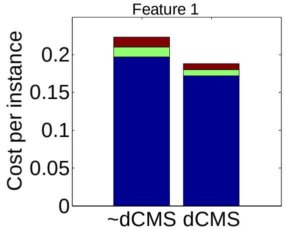
Final classification decision:

Pick the class corresponding to the score furthest away from it's corresponding optimum threshold

Results

Synthetic data: MLP neural network





MIT-DARPA intrusion detection [3]

41 features

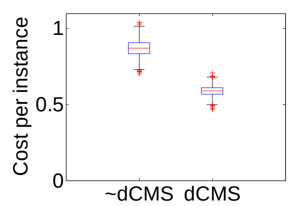
3 feature sets: traffic, content, intrinsic features

0.3 million data points

5 classes: DenialOfService, Probe, UserToRoot, RootToLocal, Normal

Ensemble of classifiers based classification system: Learn++ [4] (can perform data fusion)

True /Predicted	DOS	Probe	U2R	R2L	N
DOS	0	1	2	2	2
Probe	2	0	2	2	1
U2R	2	2	0	2	3
R2L	2	2	2	0	4
N	2	1	2	2	0



PCA reduced intrusion detection

Cost per instance	~dCMS	dCMS
Mahalanobis	1.83 $\pm$ 0.02	1.48 $\pm$ 0.01
KNN	1.32 $\pm$ 0.01	1.11 $\pm$ 0.01
Learn++ [4]	1.24 $\pm$ 0.01	0.97 $\pm$ 0.01

Other applications [5]

Cost per instance	~dCMS	dCMS
Volatile Organic Compounds	0.102 $\pm$ 0.002	0.089 $\pm$ 0.001
Optical Character Recognition	0.269 $\pm$ 0.003	0.201 $\pm$ 0.001

References:

[1] N. Lachiche and P. Flach. Improving accuracy and cost of two-class and multi-class probabilistic classifiers using ROC curves. ICML, 2003.

[2] Y. Ma and X. Ding. Robust real-time face detection based on cost-sensitive AdaBoost method. ICME, 2003

[3] The UCI KDD Archive, Information and Computer Science, University of California, Irvine, http://kdd.ics.uci.edu/ databases/kddcup99/kddcup99.html

[4] D. Parikh and R. Polikar. An Ensemble-Based Incremental Learning Approach to Data Fusion. In IEEE Transactions on Systems, Man and Cybernetics, 2007.

[5] C. Blake and C. Merz. UCI Repository of Machine Learning Database at Irvine CA, 2005. http://mlearn.ics.uci.edu/MLRepository.html