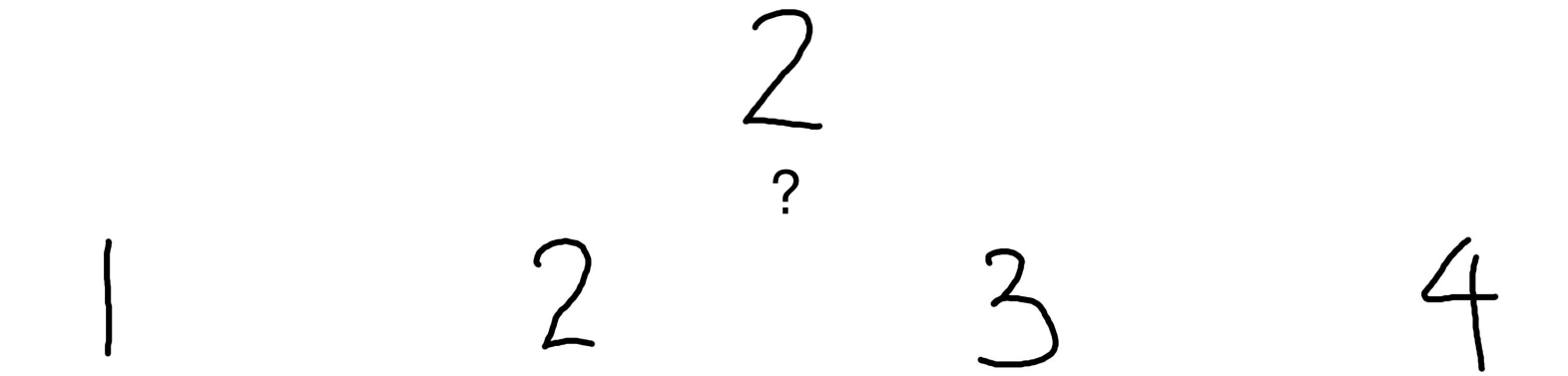
Cost-sensitive Classification: Status and Beyond

Hsuan-Tien Lin

Department of Computer Science and Information Engineering
National Taiwan University

Talk in MLRT@TAAI, 11/18/2010

Which Digit Did You Write?



- **classification**: a classic problem in machine learning

how to evaluate classification performance?

Mis-prediction Costs ($g(\mathbf{x}) \approx f(\mathbf{x})$?)

?

- ZIP code recognition:
  1: wrong; 2: right; 3: wrong; 4: wrong
- check value recognition:
  1: one-dollar mistake; 2: no mistake;
  3: one-dollar mistake; 4: **two**-dollar mistake

different applications:
evaluate mis-predictions differently

Check Value Recognition

?

1: one-dollar mistake; 2: no mistake;
3: one-dollar mistake; 4: **two**-dollar mistake

- **cost-sensitive** classification problem:
  different costs for different mis-predictions
- e.g. prediction error of g on some $(\mathbf{x}, y)$:

$$\text{absolute cost} = |y - g(\mathbf{x})|$$

cost-sensitive (as opposed to cost-less) classification:
relatively **new**, need more research

What is the Status of the Patient?

?

H1N1-infected

cold-infected

healthy

- another **classification** problem
  —grouping "patients" into different "status"

are all mis-prediction costs equal?

Patient Status Prediction

error measure = society cost

actual \ predicted	H1N1	cold	healthy
H1N1	0	1000	**100000**
cold	100	0	3000
healthy	100	30	0

- H1N1 mis-predicted as healthy: **very high cost**
- cold mis-predicted as healthy: high cost
- cold correctly predicted as cold: no cost

human doctors consider costs of decision;
can computer-aided diagnosis do the same?

Cost-sensitive Classification Setup

Given

N examples, each (input $\mathbf{x}_n$, label y_n) $\in \mathcal{X} \times \{1, 2, \ldots, K\}$
and a K by K **cost matrix** C

- $K = 2$: binary; $K > 2$**: multiclass**

Goal

a classifier $g(\mathbf{x})$ that pays a small cost $C(y, g(\mathbf{x}))$ on future **unseen** example $(\mathbf{x}, y)$

cost-sensitive classification:
a **powerful** and general setup

A Quick Overview of Selected Algorithms

cost-sensitive classification via

- relabeling
- reweighting
- relabeling + reweighting (our work, among others)
- reducing to binary classification (our work, among others)
- reducing to regression (our work)

Cost-sensitive Classification via Relabeling

(Domingos, KDD, 1999)

key idea

cost-sensitive classification
= cost-less classification + relabeling some examples based on cost

- **general** and makes any cost-less approach cost-sensitive
- but heuristic: relabel using posterior probability estimate

theoretically sound approach?

Cost-sensitive Classification via Reweighting

(Elkan, IJCAI, 2001)

key idea

cost-sensitive classification
= cost-less classification + emphasizing some costly examples

- simple and **theoretically sound**
- but applies to only binary cost-sensitive classification
  —multiclass case more complicated

theoretically sound approach
for multiclass cost-sensitive classification?

Cost-sensitive Classification via Relabeling + Reweighting (Abe et al., KDD, 2004; Lin, Caltech, 2008)

key idea

cost-sensitive classification
= cost-less classification + emphasizing and relabeling some examples

- **theoretically sound** for **multiclass**:
  good cost-less classification $\Rightarrow$ good cost-sensitive classification
- but introduces relabeling noise to the learning process
  —bad practical performance

theoretically sound approach
for multiclass cost-sensitive classification
with promising practical performance?

Cost-sensitive Classification via Pairwise Binary Classification (Beygelzimer et al, ICDM, 2003; Lin, NTU, 2010)

key idea

cost-sensitive classification
= binary classification + "Which of the two classes is of smaller cost?"

- **theoretically sound**:
  good binary classification $\Rightarrow$ good cost-sensitive classification
- **promising practical performance** (with a good binary classifier)
- does not scale well with K, the number of classes

theoretically sound approach
for large-K multiclass cost-sensitive classification
with promising practical performance?

Cost-sensitive Classification via Regression

(Tu and Lin, ICML, 2010)

key idea

cost-sensitive classification
= regression + "What is the estimated cost of each class?"

- **theoretically sound**:
  good regression $\Rightarrow$ good cost-sensitive classification
- **promising practical performance** (with a good regressor)
- **scales better with** K

what next?

Key Remaining Question: Application

theory: well-understood
algorithm: sufficiently many
application: **where? more?**

- Where does cost come from?
  - user-provided: but may not be feasible
    —consider cost intervals instead? (Liu and Zhou, KDD, 2010)
  - parameter-to-be-tuned: but currently lacks guidelines to users
    —link cost-sensitive to the true application needs?
- What are important public benchmarks?
  - semi-artificial (traditional): assigning arbitrary costs to existing sets
  - vision data with a class hierarchy?—ongoing but highly depends on feature extraction rather than costs
  - NELL data?—cost as soft-constraints
  - special types of learning problems (e.g. ranking)? others?

Assisting the users on **true application needs**
will drive future cost-sensitive classification research.

Thank you. Questions?