Building a better working world

Accurate, Data-Efficient Learning from Noisy, Choice-Based Labels for Inherent Risk Scoring

W. Ronny Huang*
Ernst & Young LLP
New York, NY
ronny.huang@ey.com

Miguel A. Perez*
Ernst & Young LLP
New York, NY
miguel.a.perez@ey.com

Application: Inherent risk scoring is used in anti-money laundering to determine riskiness of an entity *before* fraudulent acts occur

Problem: Data is scarce and the opinions of financial crime investigators are inconsistent. It is difficult to assign a risk score on an absolute scale

Hypothesis: We can use experts' choice-based feedback to determine the true label

From choice to risk score

Key point: Obtain *absolute* information from about 5-10x the amount of *relative* information

Collected choice data encoding

$$c(y_i|l) = \begin{cases} 1, & \text{if } y_i = \max_{j\in S_k} y^j \\ -1, & \text{if } y_i = \min_{j\in S_k} y^j \\ 0, & \text{otherwise} \end{cases}$$

Average to get *mean choice*

$$\bar{c}_i(y_i) = \frac{1}{q}\sum_{l=1}^{q} c(y_i|l)$$

Derive the *expected choice*

$$\langle c(y_i)\rangle = \left(\int_{-\infty}^{y_i} f(y')\,\mathrm{d}y'\right)^{s-1} - \left(\int_{y_i}^{\infty} f(y')\,\mathrm{d}y'\right)^{s-1}$$

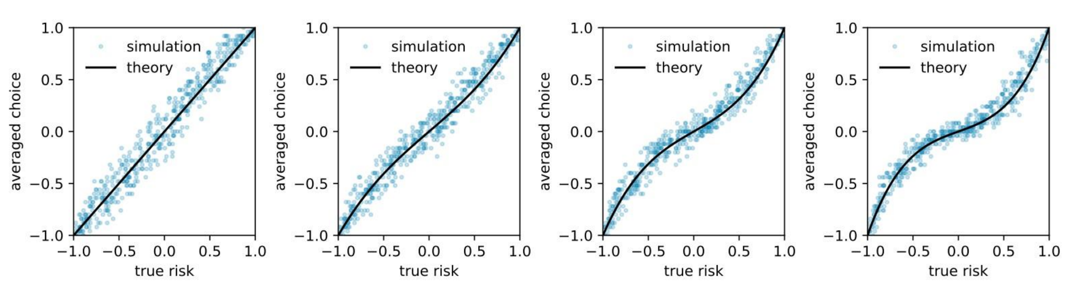
Expected choice derivation

$$\begin{aligned}\langle c(y_i)\rangle &= \mathbb{E}_{y_i\sim Y}\; c(y_i) \\ &= +1\times P(y_i = \max_{j\in S_k} y^j) \\ &\quad -1\times P(y_i = \min_{j\in S_k} y^j) \\ &\quad +0\times P(y_i \text{ is neither max nor min})\end{aligned}$$

$$\begin{aligned}P(y_i = \max_{j\in S_k} y^j) &= \prod_{j\in S_k, j\neq i} P(y_i > y^j) \\ &= \prod_{j\in S_k, j\neq i}\left(\int_{-\infty}^{y_i} f(y^j)\,\mathrm{d}y^j\right) \\ &= \left(\int_{-\infty}^{y_i} f(y')\,\mathrm{d}y'\right)^{s-1}\end{aligned}$$

Results based on a uniform label prior

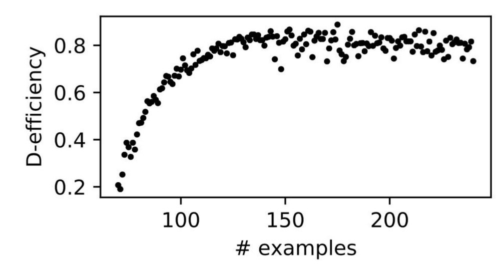
$$f(y) = 1/2 \text{ for } y \in [-1, 1]$$



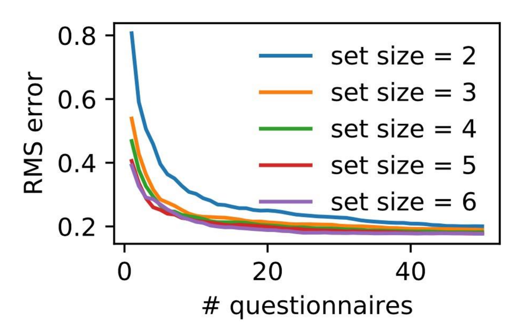
(a) Choice set size = 3 (b) Choice set size = 4 (c) Choice set size = 5 (d) Choice set size = 6

Implementation

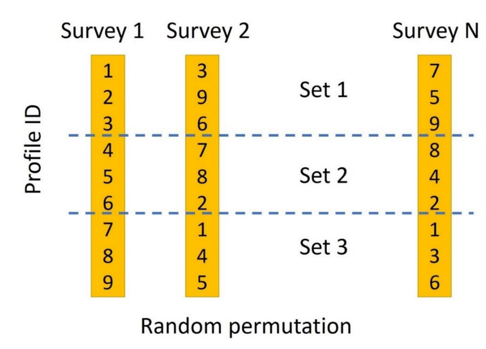
D-optimal synthetic dataset based on real schema



Error analysis helps to determine the optimal choice set size



Choice set construction by random permutation

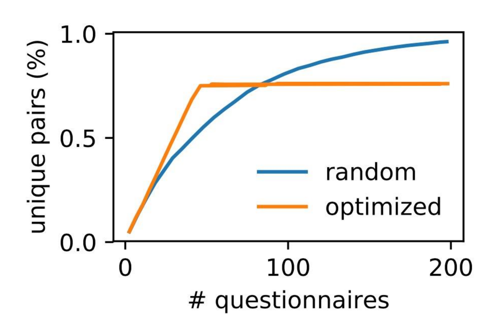


Group theory enables the fastest enumeration of choice set comparisons

Algorithm 1 Choice Set Diversifier - Size 4

Input: Range of D-optimal profiles R
Select $c \in \mathbb{N}$ and $c \in R$ such that $c = 4p$ where p is prime; p is the number of choice sets
Select three different prime numbers p_1, p_2, p_3 such that $3 < p_i$ and $p < p_i p_j$ for $i \neq j$.
Without replacement, randomly **Assign** each of the $4p$ profiles a profile key from 1 to $4p$ and place them into equal sized lists: $[A, B, C, D]$
Construct a $rank(p)$ square identity matrix, $\mathbb{I}$
For each prime number p_1, p_2, p_3:
  Construct a $rank(p)$ square permutation matrix, $\hat{G}_{p_1}, \hat{G}_{p_2}, \hat{G}_{p_3}$:
  Where elements of $[B, C, D]$ are mapped by position index i in each list such that
  $i \mapsto i + p_1$ modulo p **OR** $i \mapsto i + p_2$ modulo p **OR** $i \mapsto i + p_3$ modulo p
Define a Group Action $\hat{\mathbf{G}}$ on the vector space of ordered tuples $\langle a, b, c, d\rangle$ where
  $a \in A, b \in B, c \in C, d \in D$ such that $\hat{G} = \mathbb{I} \oplus \hat{G}_{p_1} \oplus \hat{G}_{p_2} \oplus \hat{G}_{p_3}$ and $\hat{G}^{(p+1)} = \hat{G}$
Action on a Choice Set is given by
  $\hat{G}(\langle a_1, b_1, c_1, d_1\rangle) = \mathbb{I} * a_1 \oplus \hat{G}_{p_1} * b_1 \oplus \hat{G}_{p_2} * c_1 \oplus \hat{G}_{p_3} * d_1 = \langle a_1, b_{1+p_1}, c_{1+p_2}, d_{1+p_3}\rangle$
Append each ordered list $[A, B, C, D]$ so that $U = A + B + C + D$. U is the first questionnaire
Apply $\hat{\mathbf{G}}$ to the randomized ordered list U to generate the next questionnaire
Stop when $\hat{G}$ has been applied to U p number of times
Return p number of unique questionnaires $U, U', U'', \cdots$

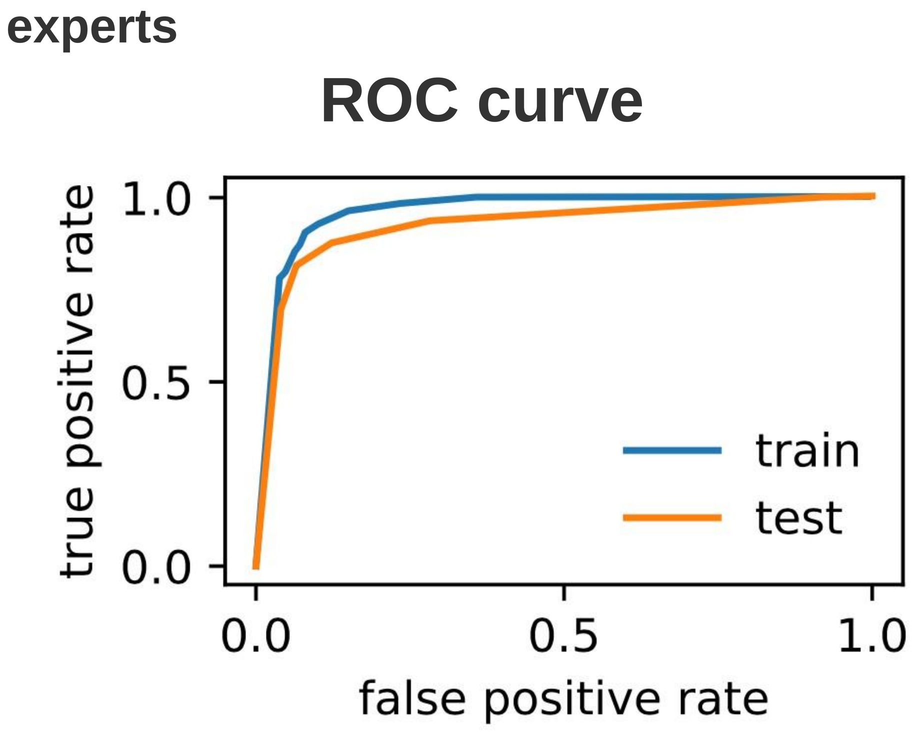
Choice set diversifier algorithm allows for fewer expert evaluations



Results

Model performance and test results at a large financial institution

Synthetic examples evaluated by real experts



Metric	Train	Test
AUC	97%	93%
Classification Error	8%	11%

Real world examples

Population Group	Profiles	SMA Escalations	Escalation Rate
IRM Selected Alerted Profiles	1,500	28	1.87
Remaining Scenario Alerted Profiles	2,500	3	0.12

A choice-based questionnaire for determining true labels

```
SET 1 of 25:
         age  country  pep     years
0  twenties  nigeria   no  twothree
1  thirties  nigeria  yes       one
2   fifties    italy  yes  twothree
3   sixties  nigeria  yes  fourmore
Enter indices of the most/least
risky customer (0|1|2|3): 1,0
```

(a) Example of a choice set.

		age	country	pep	years	Encoded choice c
Set 1	0	twenties	nigeria	no	twothree	-1.0
	1	thirties	nigeria	yes	one	1.0
	2	fifties	italy	yes	twothree	0.0
	3	sixties	nigeria	yes	fourmore	0.0
Set 2	4	thirties	italy	no	one	0.0
	5	thirties	nigeria	no	one	0.0
	6	thirties	usa	yes	one	1.0
	7	forties	usa	no	one	-1.0
	..	...	...	...	...	...

(b) Example data layout.

* Equal contribution