Distributed Data Mining in Credit Card Fraud Detection

Large scale data mining is used in an attempt to improve upon the state of the art in commercial credit card transaction safety practices. Efficient fraud detectors can be garnered from massive data sets, but timely and efficient data mining techniques must be utilized. The current research evaluates several proposed data mining and cleaning techniques.

Background:

- Credit card transactions represent the majority of in-person transactions, and nearly 100 percent of online transactions.
- Credit card data is extremely valuable to hackers and thieves alike.
  - -Individual credit cards may be stolen, or their information physically taken, and used by low level fraudsters
  - -Technically sophisticated hackers can seize thousands of credit card numbers simultaneously, selling them on the black market in bundles for a huge profit. These perpetrators, the market system they use, and their buyers, are often quite good at covering their tracks.
- Tools used to detect fraud through data mining must meet several criterion:
  - -**Efficacy**: The technique must effectively detect fraud.
  - -**Scalability**: The technique must be capable of being used throughout a credit network.
  - -**Efficiency**: The technique must be computable in a time and resource constrained environment.
  - -**Uniformity/Validity**: The technique must detect fraud in an often skewed and uncorrected data environment.

Findings:

- Using a cost model, multiple learned fraud detectors are combined and deemed as theoretically useful.
- Empirical results demonstrate significantly reduced loss due to fraud through distributing the data mining fraud model below:

- Given: $(x_1, c_1, y_1), \ldots, (x_m, c_m, y_m); x_i \in \mathcal{X}, c_i \in \mathbb{R}^+, y_i \in \{-1, +1\}$
- Initialize $D_1(i)$ (such as $D_1(i) = c_i / \sum_j^m c_j$)
- For $t = 1, \ldots, T$:
  1. Train weak learner using distribution D_t.
  2. Compute weak hypothesis $h_t : \mathcal{X} \to \mathbb{R}$.
  3. Choose $\alpha_t \in \mathbb{R}$ and $\beta(i) \in \mathbb{R}^+$.
  4. Update

$$D_{t+1}(i) = \frac{D_t(i)\exp\Big(- \alpha_t y_i h_t(x_i) \boxed{\beta\Big(\text{sign}(y_i h_t(x_i), c_i)\Big)}\Big)}{Z_t}$$

  where $\beta(\text{sign}(y_i h_t(x_i)), c_i)$ is a cost-adjustment function. Z_t is a normalization factor chosen so that D_{t+1} will be a distribution.

- Output the final hypothesis:

$$H(x) = \text{sign}(f(x)) \quad \text{where} \quad f(x) = \left(\sum_{t=1}^{T} \alpha_t h_t(x) \right)$$

Figure 1: AdaCost

Implications:

-Additional benefits can be garnered through addressing database compatibility. In the above model, all databases are assumed to be schematically similar.

-This research focuses on large-scale fraud detection. The focus in the LP industry is often smaller scale and involves prevention or apprehension rather than detection. Understanding detection on a large scale may be a useful added tool for LP executives as they navigate the credit card fraud space.