COLUMBIA UNIVERSITY

COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

Meta-Learning Approach for Credit Card Fraud Detection

Moorissa Tjokro

Department of Computer Science, Columbia University

Topics in Financial Computing
COMS W4995
Moorissa Tjokro | *mmt2167*
moorissa.tjokro@columbia.edu
www.moorissa.com

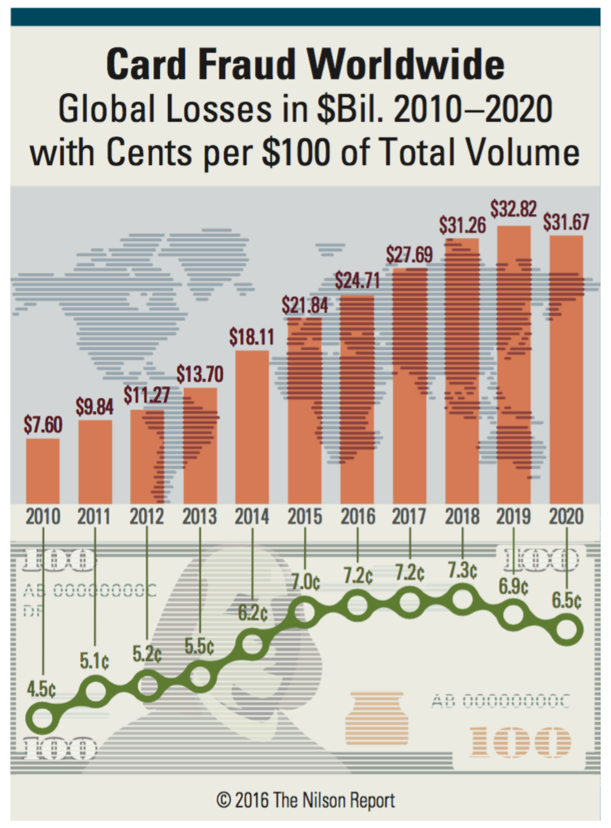
PROBLEM DESCRIPTION

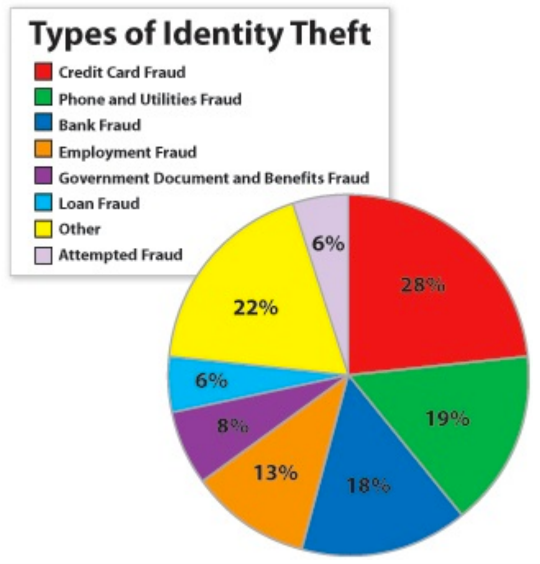
Rapid growth of technology over the past decade has created proliferation of credit card use worldwide.

In 2012, U.S. Department of Justice noted approximately nearly half, **$5.55 billion** credit card fraud, occurred in the US alone.

Major companies took actions to prevent further losses:

- Bank of America agreed to pay $16.5 billion for resolving financial fraud case
- IRS paid closer attentions to professionals and academicians who committed fraud.





With billions of dollars incurred by financial institutions every year, detecting fraudulent behaviors has been of significant importance.

This research studies the implementation of Bayesian Networks, K-Nearest Neighbors, and Decision Trees that are used in a proposed meta-learning solution for improving a Bayesian Networks single-approach for credit card fraud detection.

ADVANTAGES

Advantages and Key Design Considerations:

1. **Accuracy**
   - ✓ Several studies have shown that Bayesian learning proves to have higher accuracy in detection process compared to many common machine learning techniques such as artificial neural networks.
2. **Ability to handle noise**
   - ✓ The probabilistic nature of the Bayesian network allows for handling noise in the data.
3. **Flexibility**
   - ✓ Bayesian networks come from a family of models called graphical models—a family of flexible and interpretable models. This allows for a more suitable approach to the nature of fraudulent behaviors.
4. **Common use in a variety of environments**
   - ✓ The application of Bayesian networks is often found in various industries from mechanical engineering to medical field. Commonalities and its familiarity to many triggered the use of approach in the finance sector.

TRXID	Client ID	Date	Amount	Location	Type	Merchant Group	Fraud
1	1	2/1/12 6:00	580	Ger	Internet	Airlines	No
2	1	2/1/12 6:15	120	Eng	Present	Car Rent	No
3	2	2/1/12 8:20	12	Bel	Present	Hotel	Yes
4	1	3/1/12 4:15	60	Esp	ATM	ATM	No
5	2	3/1/12 9:18	8	Fra	Present	Retail	No
6	1	3/1/12 9:55	1210	Ita	Internet	Airlines	Yes

BAYESIAN NETWORKS APPROACH

Overview

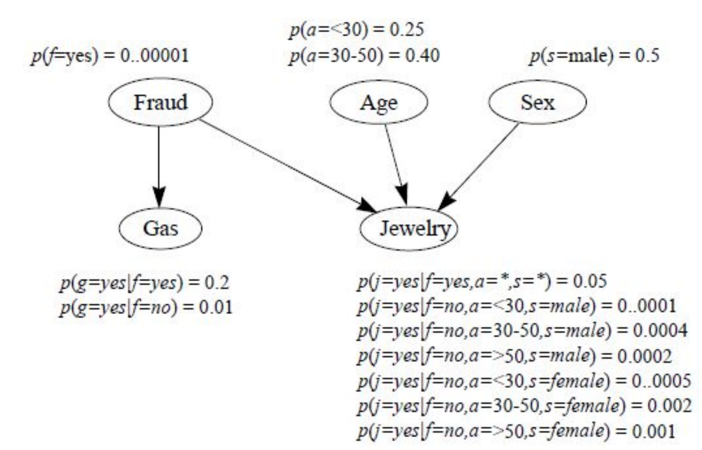
A Bayesian network is a graphical model for probabilistic relationships among a set of variables. Each variable has a finite set of mutually exclusive states. A set of directed links or arrows connects pairs of nodes. Intuitively, an arrow from node X to node Y means that X has a direct influence on Y.

- ❖ Probabilistic terms between predictors and outcome variables (Luttrell, 1994).
- ❖ Probability theory, mathematical statistics (Sox et al., 1994).

Overall, Bayesian Networks provide a comprehensive method of representing relationships and influences among nodes (variables).

Methodology

A Bayesian network for detecting credit-card fraud



Direction of arcs:
from parent to descendant node

Parents of node X_i: Pa_i

Pa(Jewelry) = {Fraud, Age, Sex}

Advantages

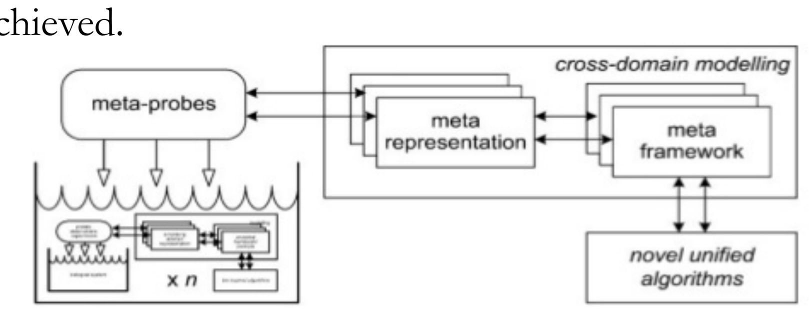
- ❖ Don't need complete dataset
- ❖ Can learn causal relationships
- ❖ Combines domain knowledge and data
- ❖ Avoids overfitting – don't need test data

Disadvantages

- ❖ Limited pattern recognition: can only exploit causal influences that are recognized by the programmer.
- ❖ Requires a comparatively large amount of effort.

META-LEARNING APPROACH

This meta-classifier model consists of **3 base classifiers** constructed using the decision tree, k-nearest neighbor algorithms, and Bayesian network. The naïve Bayesian algorithm was used particularly as the meta-level algorithm to combine the base classifier predictions to produce the final classifier. Results from the research show that when a meta-classifier was deployed in series with the Bank's existing fraud detection algorithm improvements of up to **28%** to their existing system can be achieved.



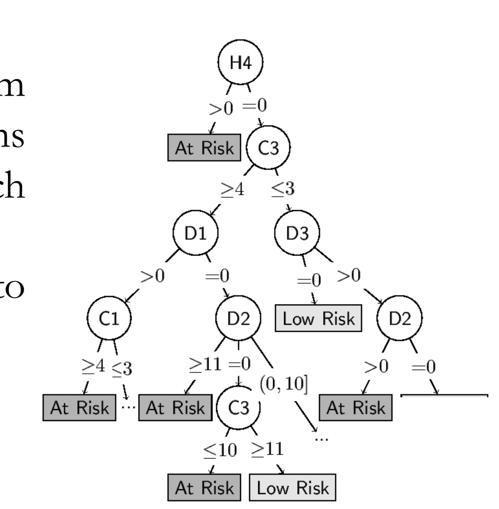
The meta-learning technique aims to coalesce the results of multiple learners to improve prediction accuracy and to utilize the strengths of one method to complement the weaknesses of another.

Therefore, the combiner strategy is used, in which the attributes and correct classifications of credit card transaction instances are used to train multiple base classifiers. By combining the original attributes, the base classifier predictions, and the correct classification for each instance, anew "combined" dataset is created which is used as the training data to generate the meta-level classifier. The predictions from the meta-level classifier are then used as the final predictions in the combiner strategy.

METHODOLOGY

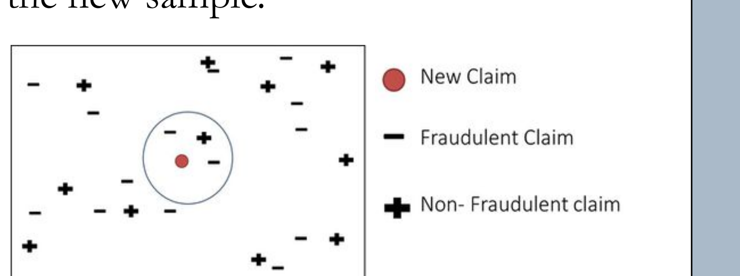
Decision Tree

- ❖ Classification rules, extracted from decision trees, IF-THEN expressions and all the tests have to succeed if each rule is to be generated.
- ❖ Separates the complex problem into many simple ones.
- ❖ Resolves the sub-problems through repeated usage.



K-Nearest Neighbor

- ❖ Locate the nearest neighbors.
- ❖ Neighbors used to classify the new sample.
- ❖ Easy detect.
- ❖ Unsupervised learning.



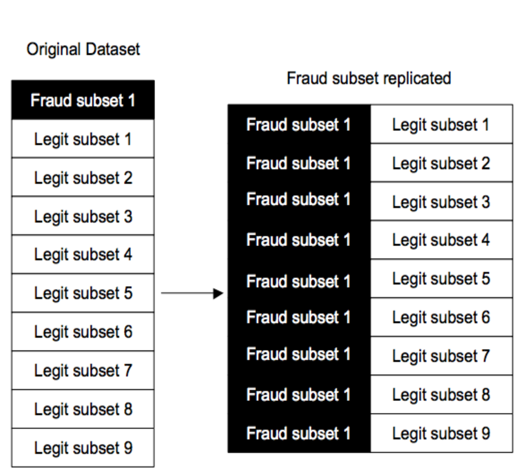


Figure 1: Constructing a 50:50 distribution for the training datasets

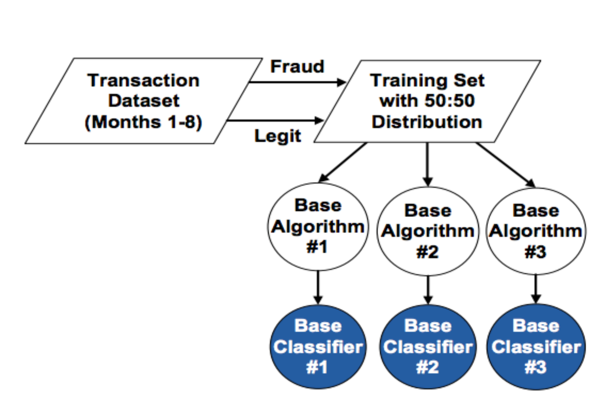


Figure 2: Stage 1 – Training stage in the Meta-learning process

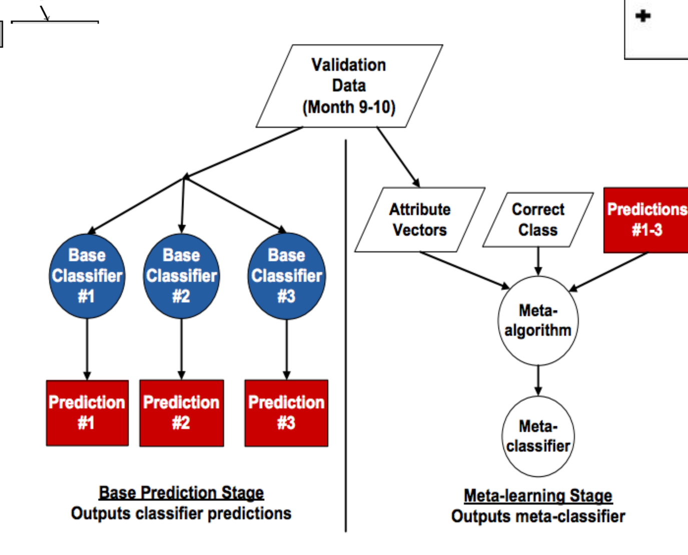


Figure 3: Stage 2 & 3 – Generating the base classifier predictions and constructing the meta-classifier

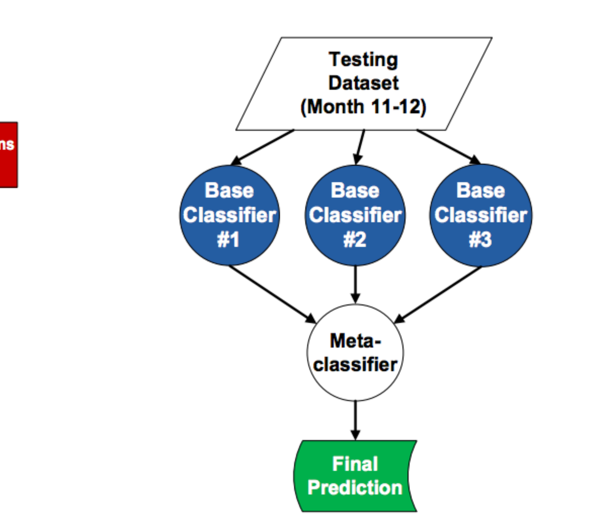


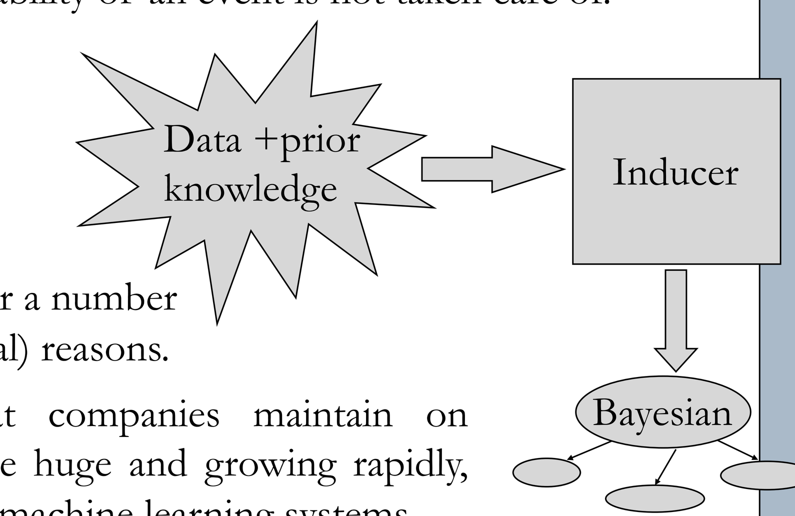
Figure 4: Stage 4 – Generating the final predictions for dataset

KEY LIMITATIONS

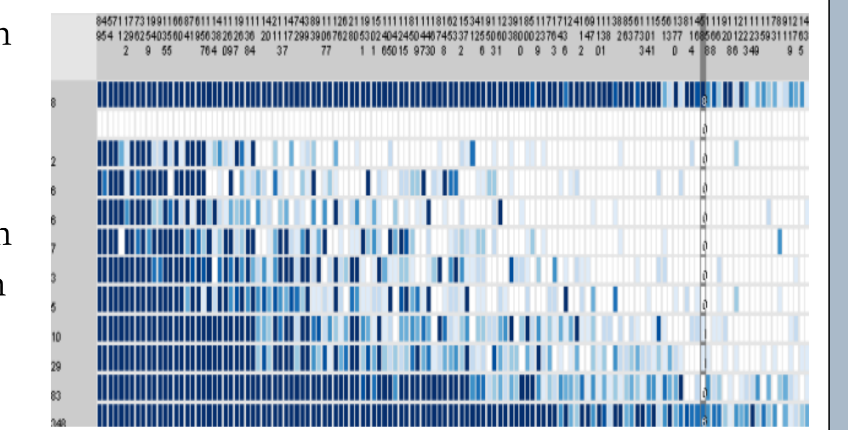
We observe some limitations of the Bayesian Networks:

- ❖ Typically require initial knowledge of many probabilities. Quality and extent of prior knowledge play an important role
- ❖ Significant computational cost (NP hard task)
- ❖ Unanticipated probability of an event is not taken care of.

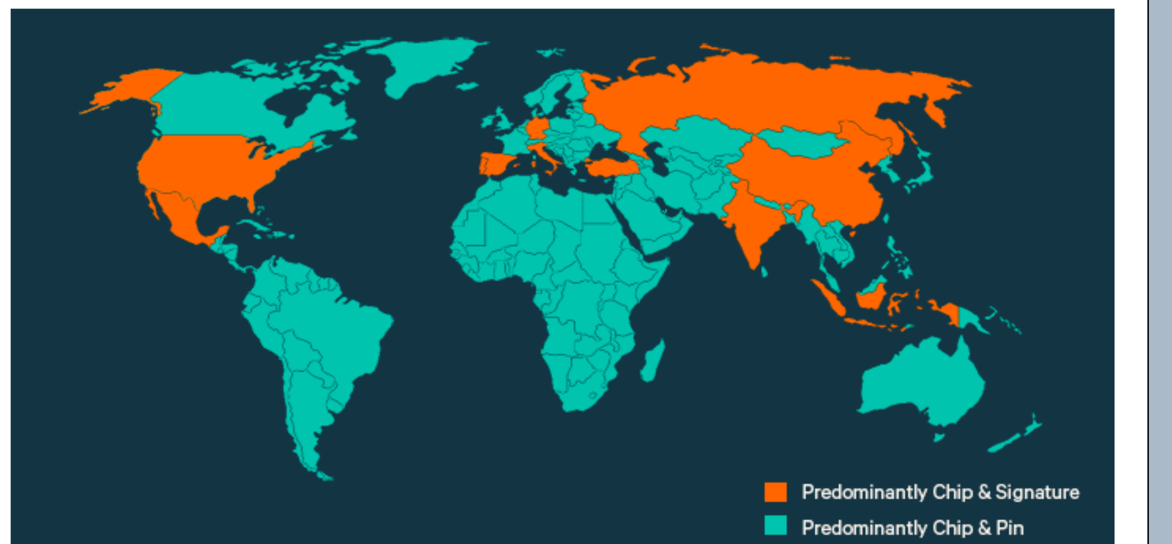
Key difficulties in building a desired system include:



- Financial companies don't share their data for a number of (competitive and legal) reasons.
- The databases that companies maintain on transaction behavior are huge and growing rapidly, which demand scalable machine learning systems.
- Real-time analysis is highly desirable to update models when new events are detected.
- Easy distribution of models in a networked environment is essential to maintain up to date detection capability.



CHALLLENGES



Challenges for the meta-learning approach:

- Necessary to perform an empirical evaluation (e.g. cross-validation) of the candidate algorithms on a problem.
- High cost for generating a whole set of meta-example – depending on the number and complexity of the candidate algorithms, the methodology of empirical evaluation and the amount of available problems.
- Certain applications of datasets may pose a unique challenge for implementations because of their high dimensionality, multiple classes, noisy data and missing values.
- Assumes that the features used to represent meta-instances are sufficiently relevant, but some features may not be directly relevant, and some features may be redundant or irrelevant.
- Predictable factors such as the available amount of training data, the spatial variability of the effective average distance between data samples, and the type and amount of noise in the data can set influence to classifiers to a significant degree.