Supercharging Decisions through Additional Data Sources

Credit Scoring & Credit Control IX

8 September 2005

Dr. Andrew Jennings
General Manager
Fair Isaac

Agenda

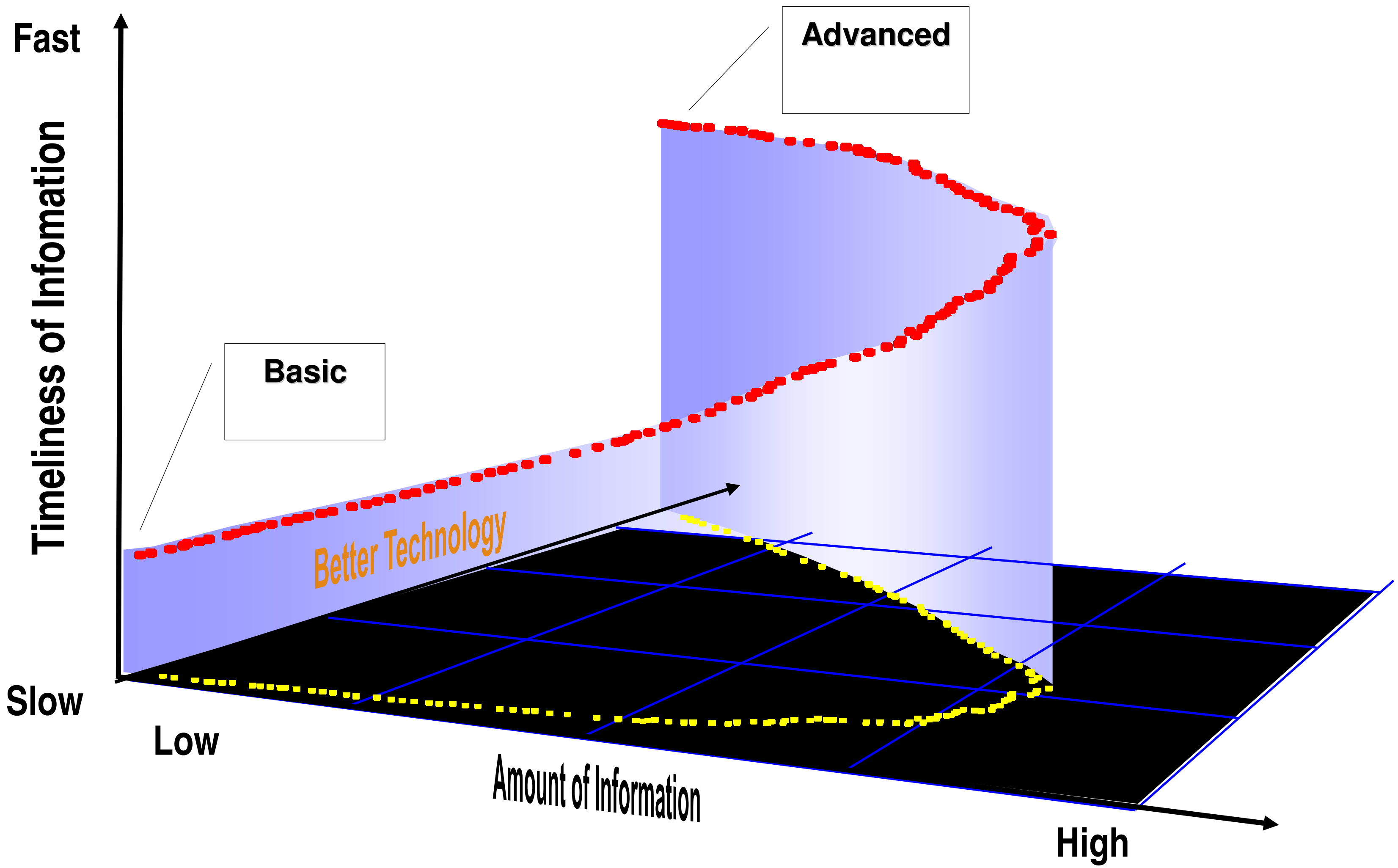
- **Motivation**
- **Solutions**
- **Challenges**
- **Case Studies**
- **Summary**

Current banking industry landscape

The problem:

- Customers getting other financial products elsewhere
- Customers poorly understood
- Daily transactions not analyzed
- Attrition indicators not recognized
- Opportunity indicators ignored
- Customers at risk of over-indebtedness

Where will the improvement come from?



Agenda

- Motivation
- Solutions
  - Advanced technology
  - Increased understanding
  - Better decisions
- Challenges
- Case Studies
- Summary

HF47

Advanced technology in high-tech scientific and military applications can be migrated to commercial use

- **Language and Text Mining from massive multi-lingual textual data**
  - Information retrieval and Knowledge extraction
  - Automatic document categorization, trend discovery, and decisioning
  - Advanced language representation, understanding, and question answering
- **Bioinformatics Research**
  - Unique DNA ‘fingerprint’ identification

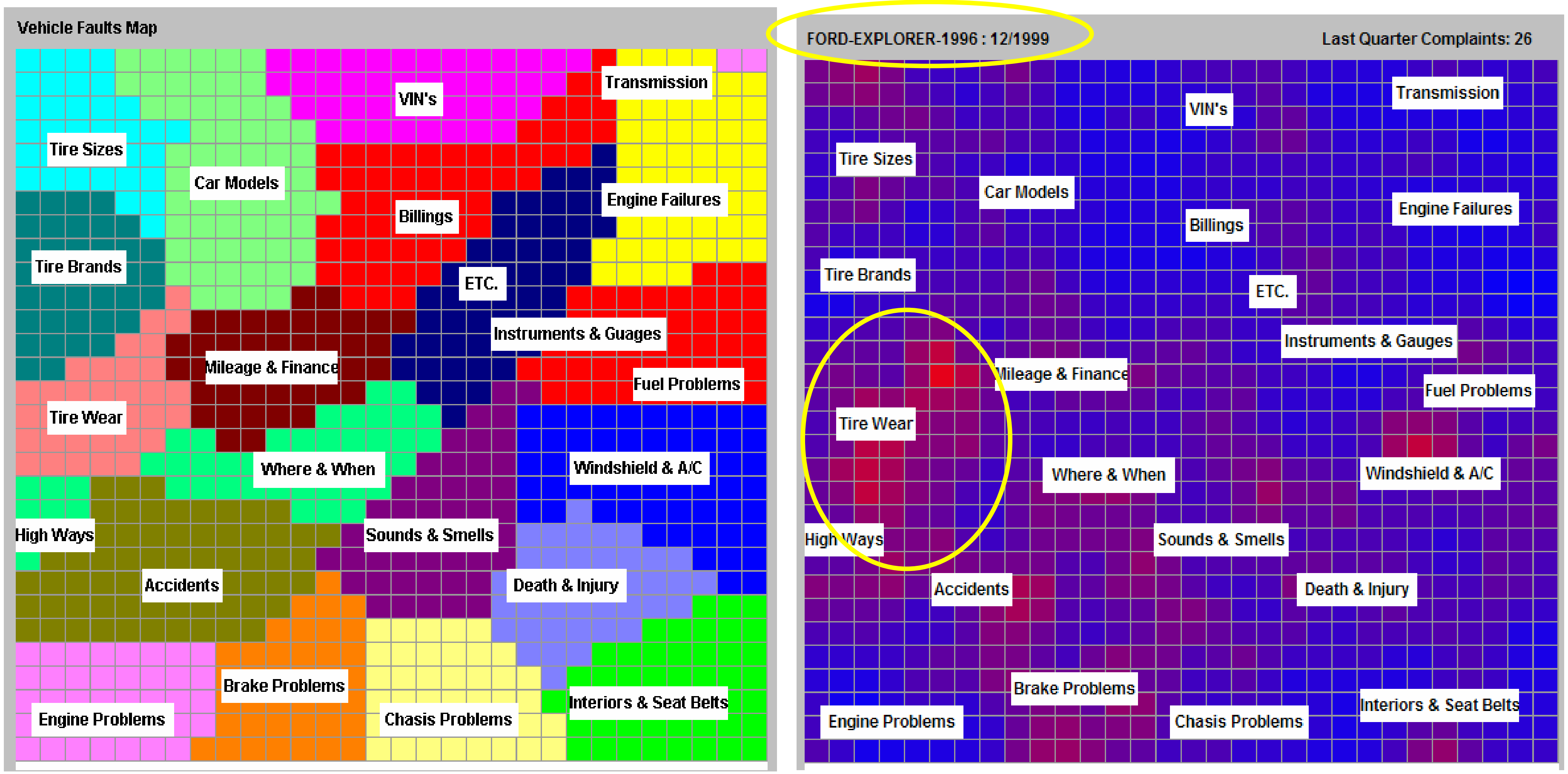
- **Advanced Military Research**
  - Military target recognition
  - Hierarchical reinforcement learning algorithms for war-game simulations and automatic strategy learning
- **Neuroscience**
  - Sparse-coded brain-like associative memories

Slide 6

HF47 Make tighter connection to what we're going to talk about - eg context vectors; peacock
Hollis Fishelson-Holstine, 24/08/2005

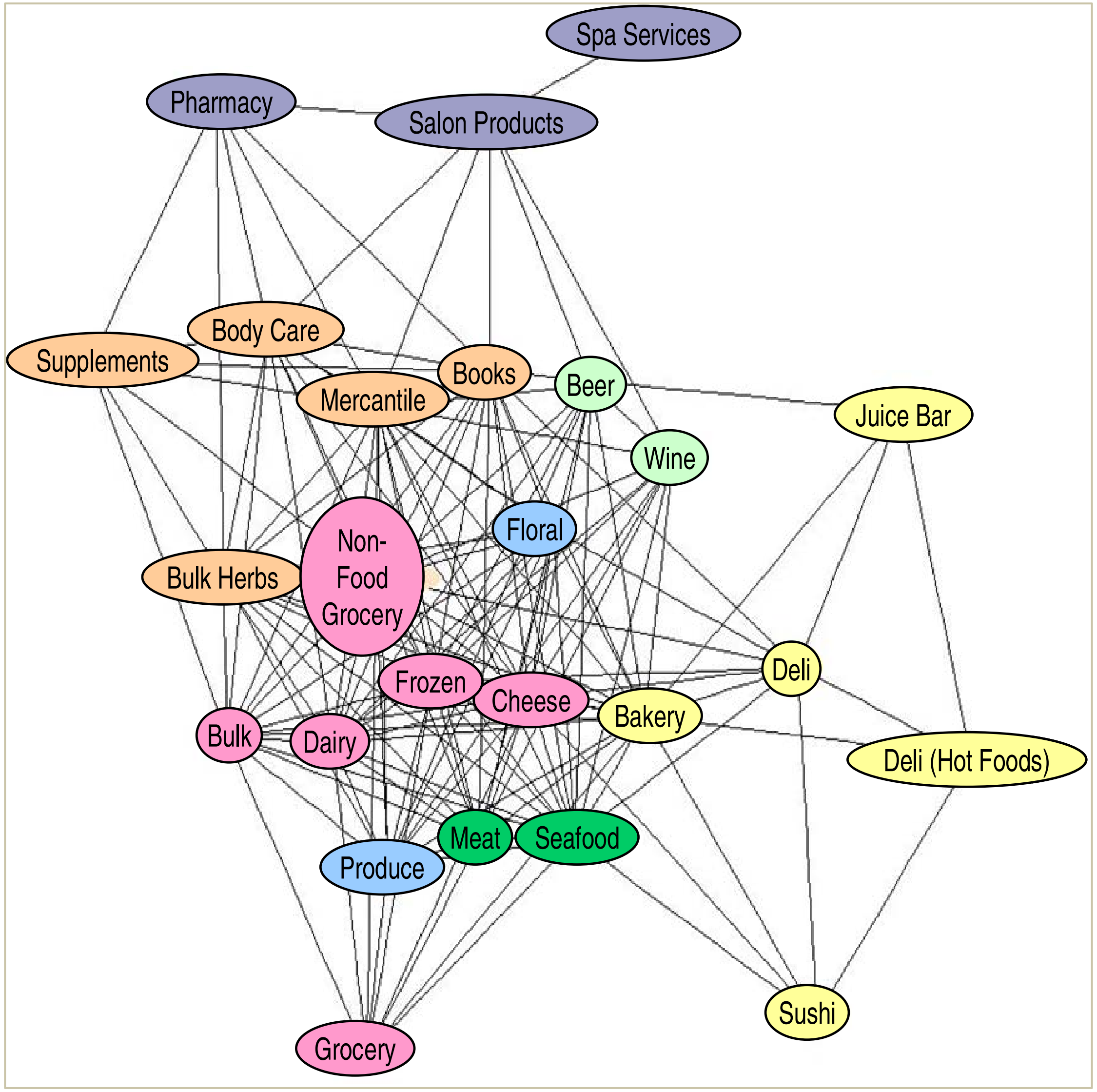
Visualizing trends in call center data

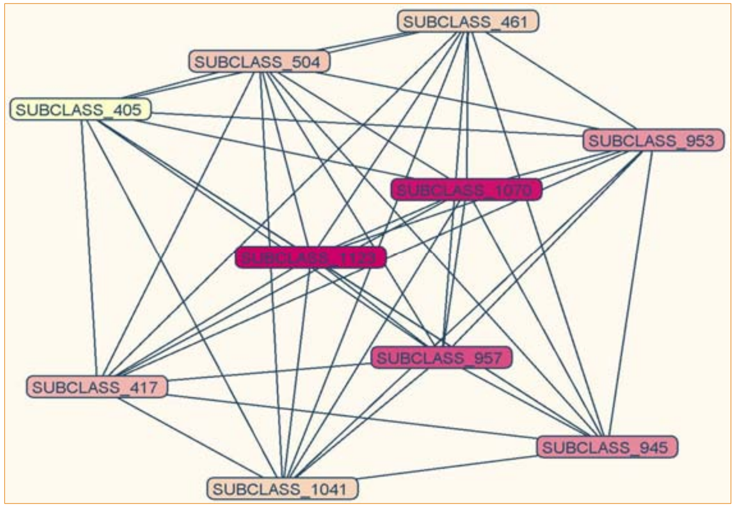
Early warning of parts failure in automotive accident data

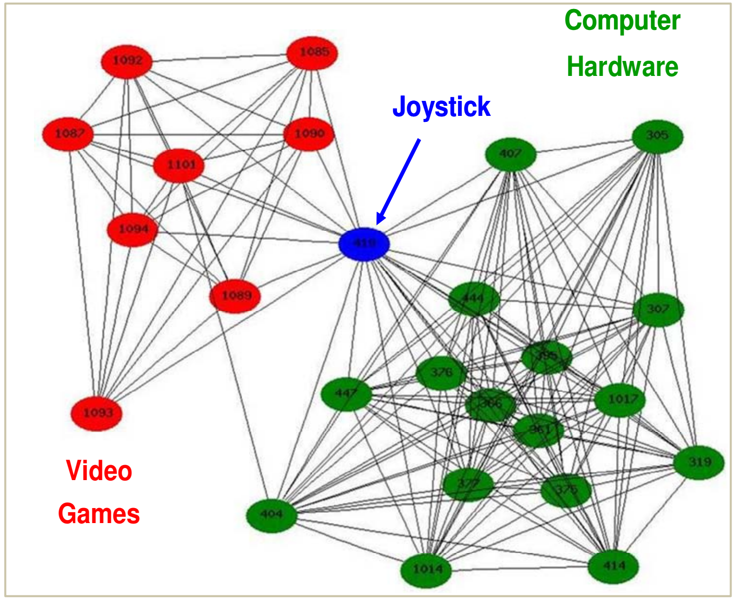


“The Ford-Firestone dispute blew up in August 2000 and is still going strong. In response to claims that their 15-inch Wilderness AT, radial ATX and ATX II tire treads were separating from the tire core--leading to grisly, spectacular crashes--**Bridgestone/Firestone** recalled 6.5 million tires, mostly original equipment on the Ford Explorer, the world's top-selling sport utility vehicle (SUV).” - Forbes

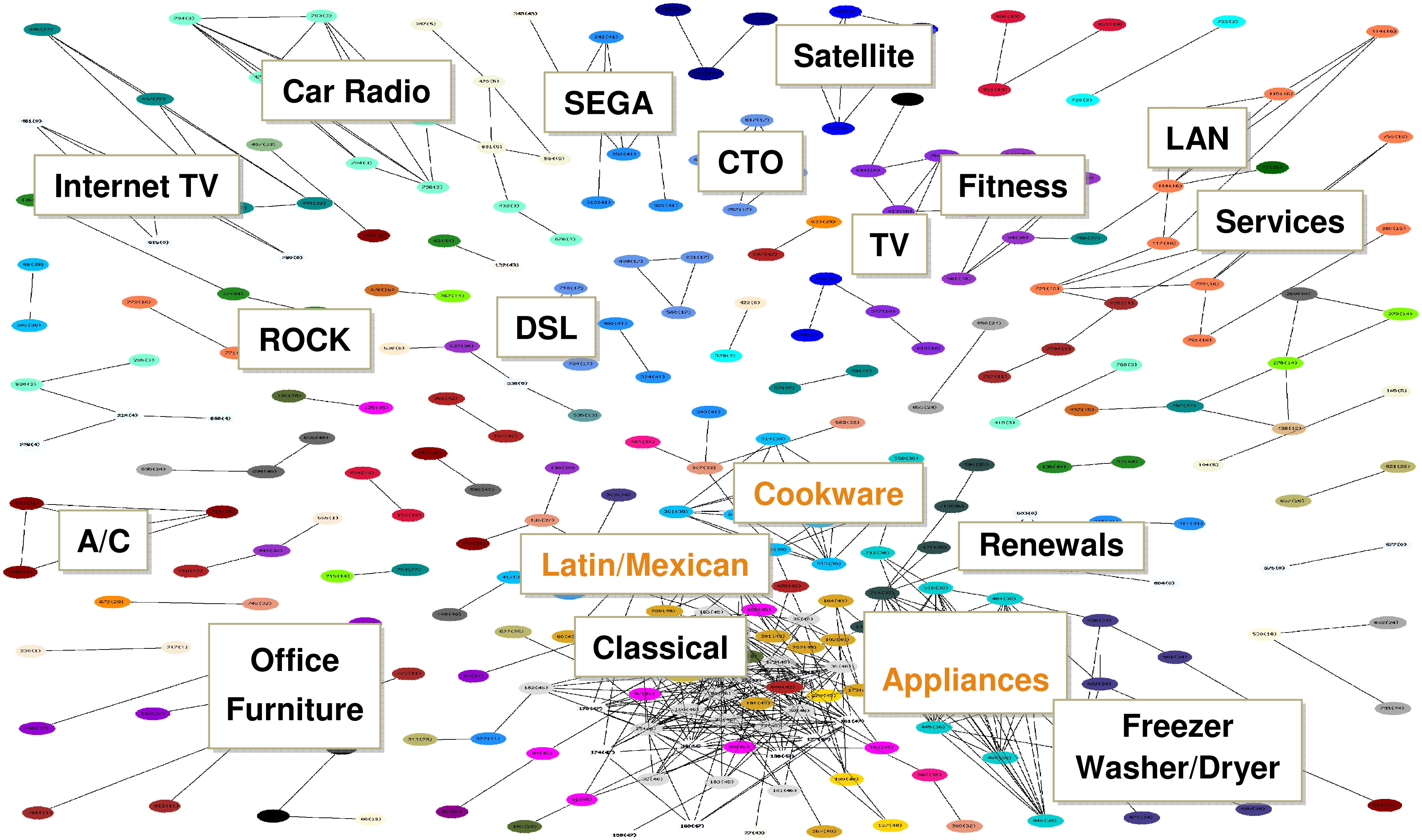
Example - Pairwise Co-occurrence Consistency (PCoC or "Peacock")







Example - Pairwise Co-occurrence Consistency (PCoC or "Peacock")



Agenda

- Motivation
- Solutions
  - Advanced technology
  - Increased understanding (additional information)
  - Better decisions
- Challenges
- Case Studies
- Summary

HF42

Notion of "Transactions"

- Transaction is a time stamped record of an event or state associated with an entity or a unique key
- Sorted and merged transactions provide a complete time series history
- Characteristics of transactions
  - Transactions are heterogeneous
  - Transaction volumes can be enormous
  - Transaction timing is irregular
  - An opportunity to change the outcome – real time decisions

Slide 11

HF42 Andy was going to maybe replace this with a picture from one of the old TJ presentations
Hollis Fishelson-Holstine, 28/08/2005

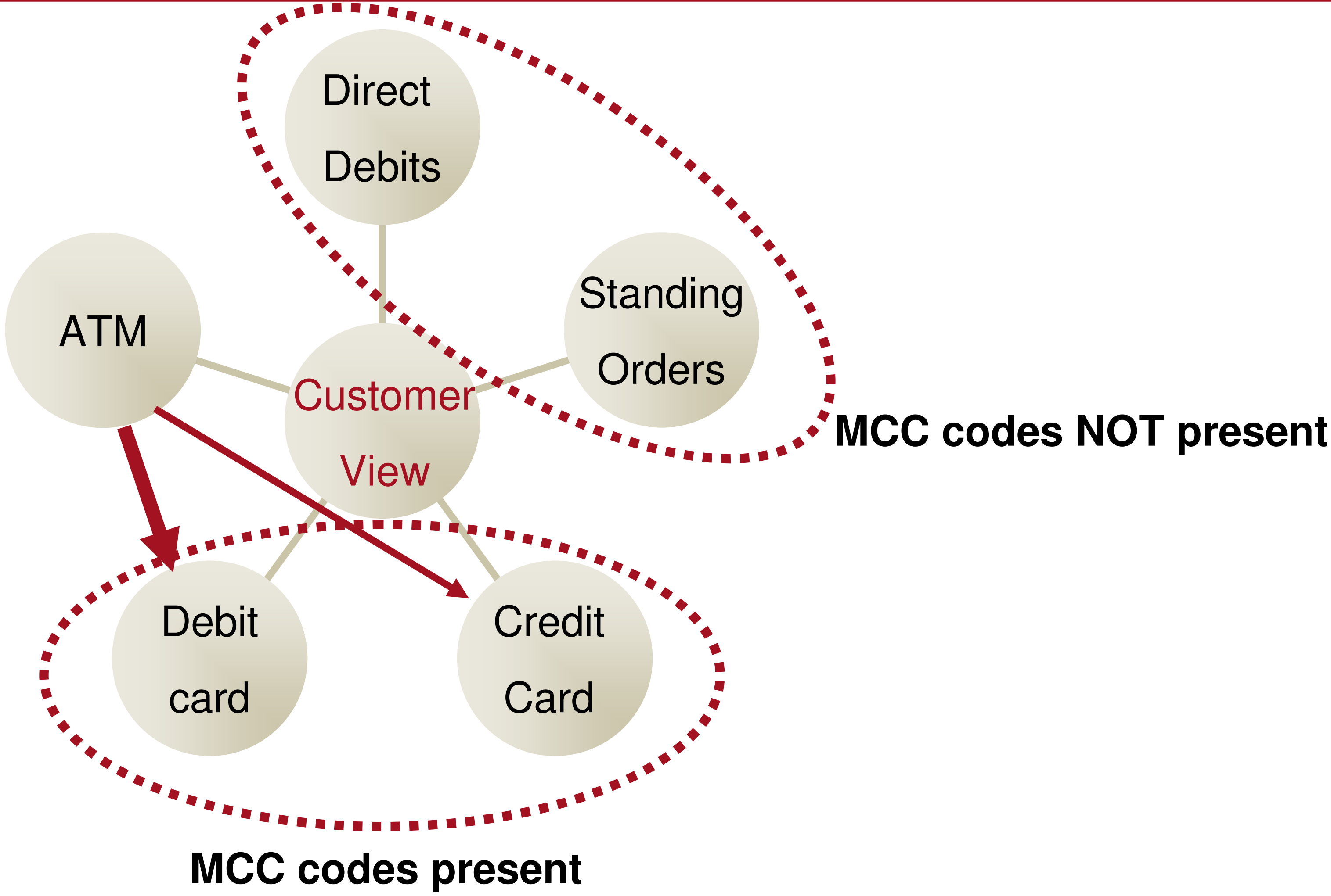
The Key to Transaction Data is to Understand Each Field

- **Who**
  - Account number, name, cardholder country, cardholder zip
- **When**
  - Transaction date and time, velocity, shopping times
- **Where**
  - Geographic region / metro classification, proximity, distance
- **What**
  - Type of business (MCC/SIC), goods sold, services offered
  - High-end / low-end, typical spend: Chain, brand, loyalty
  - Product codes (SKUs) and brand preferences
- **How much**
  - Cash usage, charge-backs, payments
  - Spending compared to others by Merchant, MCC/SIC, Post Code

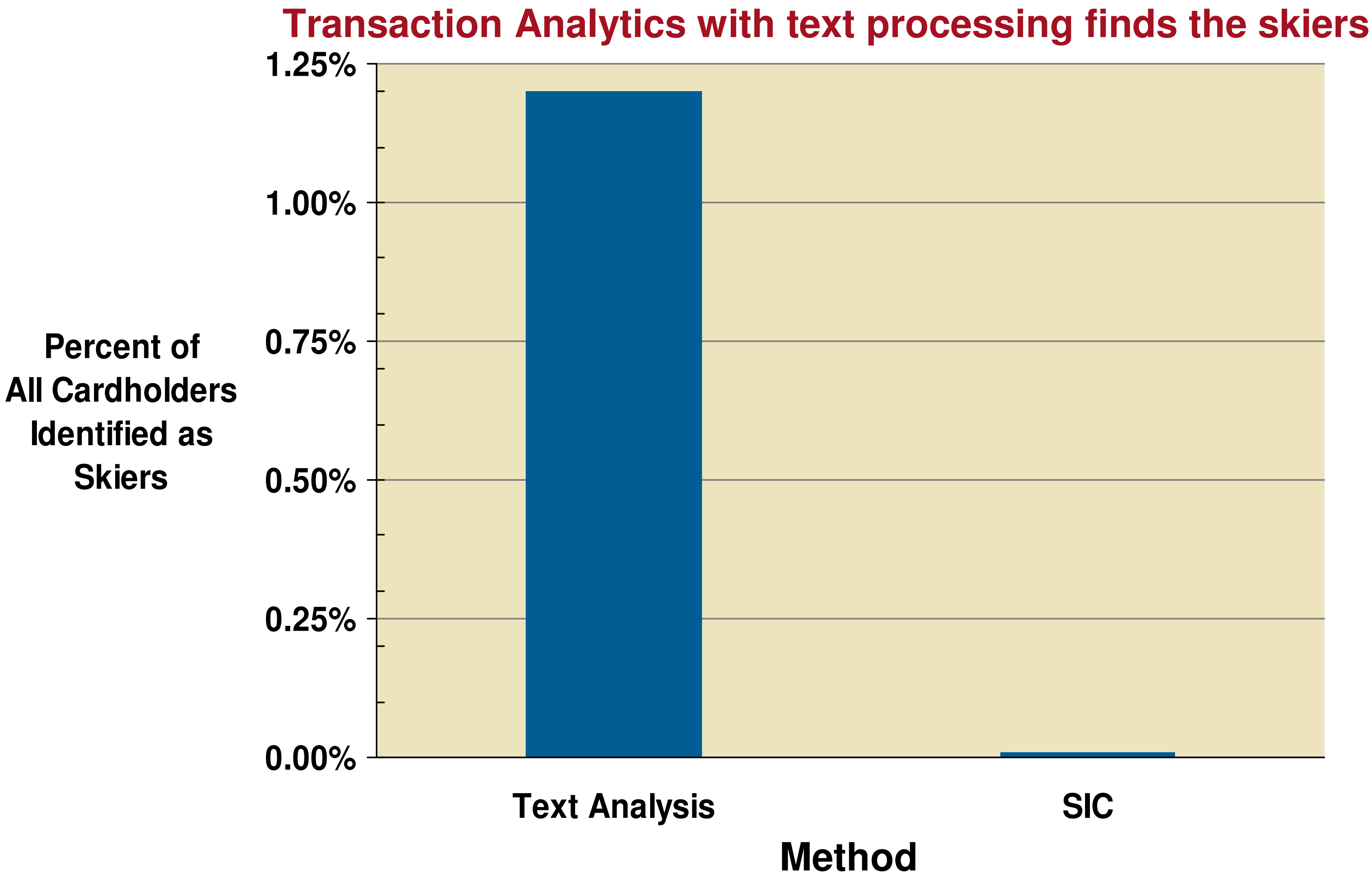
Understanding further enhanced by text

- The text that you rarely use in analytics
  - *Collectors notes, Call center notes, phone call transcripts*
  - Notes in credit bureau reports
  - Annotations on accounts, text in application forms (online/offline)
  - *Merchant names in credit card transactions*
  - Codes, annotations and other unstructured data that people understand, but is ignored in modeling & scoring.

What is the Problem with MCC Codes?



SICs simply cannot identify skiers



Example Infant Purchasers
Fair Isaac's Targeted Solution

#1

Merchant Name	Payment Method	Number of Transactions	Amount	ATV
~~MOTHERCARE WORLD~~	~~Debit Card~~	~~13~~	~~£268~~	~~£20~~
~~MOTHERCARE UK LTD~~	~~Debit Card~~	~~6~~	~~£67~~	~~£11~~
~~ADAMS CHILDRENSWEAR~~	~~Debit Card~~	~~2~~	~~£33~~	~~£16~~
BUSH BABES	Debit Card	2	£99	£49
KIDDICARE	Debit Card	2	£156	£78
BLOOMING MARVELLOU	Debit Card	1	£17	£17
~~ADAMS CHILDRENSWEA~~	~~Debit Card~~	~~1~~	~~£14~~	~~£14~~
~~MOTHERCARE COM INT~~	~~Debit Card~~	~~1~~	~~£72~~	~~£72~~
BON BLEU	Debit Card	1	£8	£8
		29	**£726**	**£25**

#2

Merchant Name	Payment Method	Number of Transactions	Amount	ATV
~~MOTHERCARE WORLD~~	~~Debit Card~~	~~15~~	~~£216~~	~~£14~~
~~MOTHERCARE UK LTD~~	~~Debit Card~~	~~5~~	~~£61~~	~~£12~~
JOKIDS LTD	Debit Card	2	£18	£9
~~ADAMS CHILDRENSWEA~~	~~Debit Card~~	~~1~~	~~£10~~	~~£10~~
		23	**£305**	**£13**

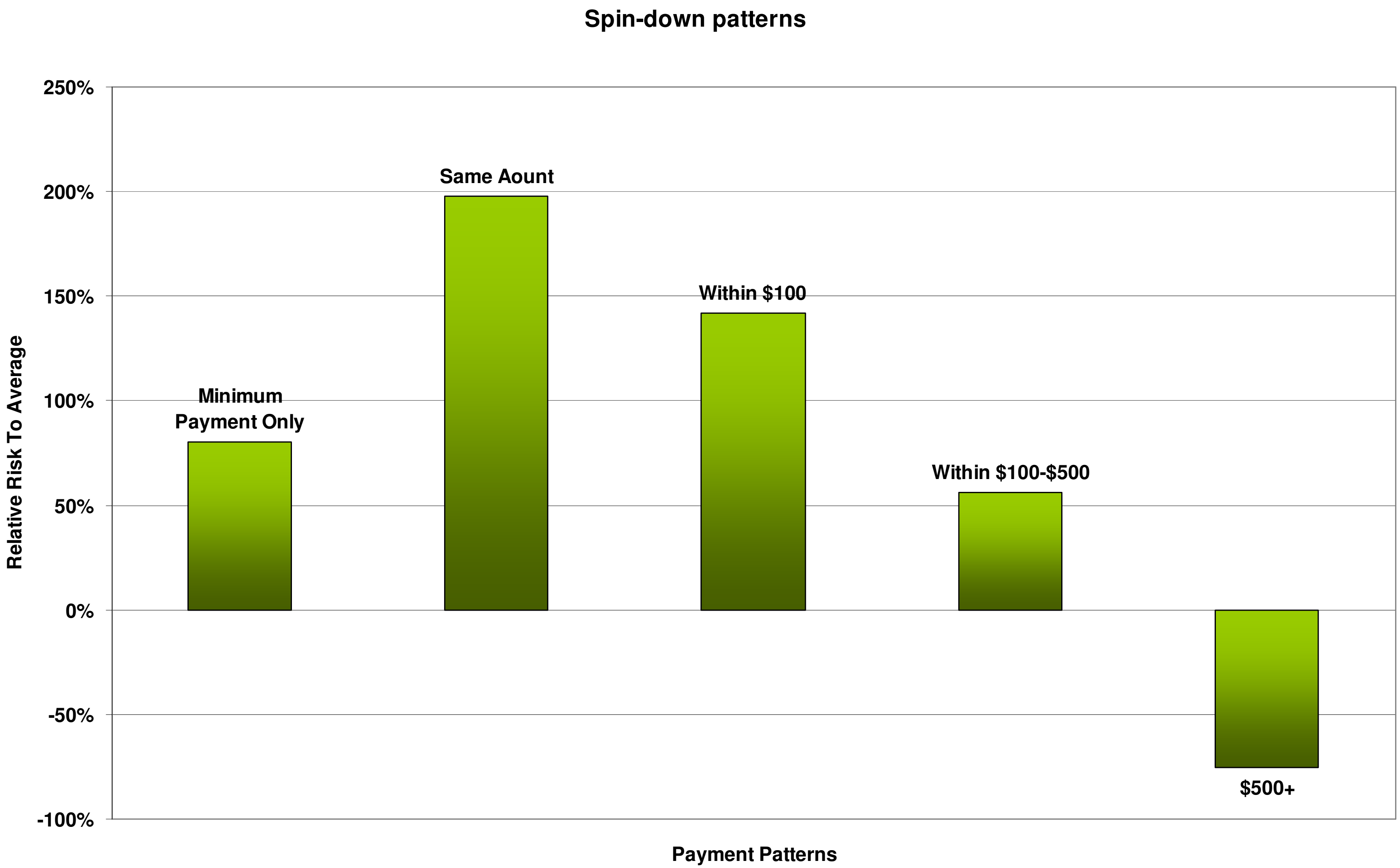
#3

Merchant Name	Payment Method	Number of Transactions	Amount	ATV
~~MOTHERCARE WORLD~~	~~Debit Card~~	~~10~~	~~£474~~	~~£47~~
NIPPERS	Debit Card	7	£360	£51
~~MOTHERCARE WORLD~~	~~Credit Card~~	~~3~~	~~£241~~	~~£80~~
~~MOTHERCARE UK LTD~~	~~Debit Card~~	~~2~~	~~£81~~	~~£40~~
~~ADAMS CHILDRENSWEAR~~	~~Debit Card~~	~~1~~	~~£22~~	~~£22~~
		23	**£1,178**	**£51**

MCC codes miss so many merchants

HF43

How much ? Balance Spin-down

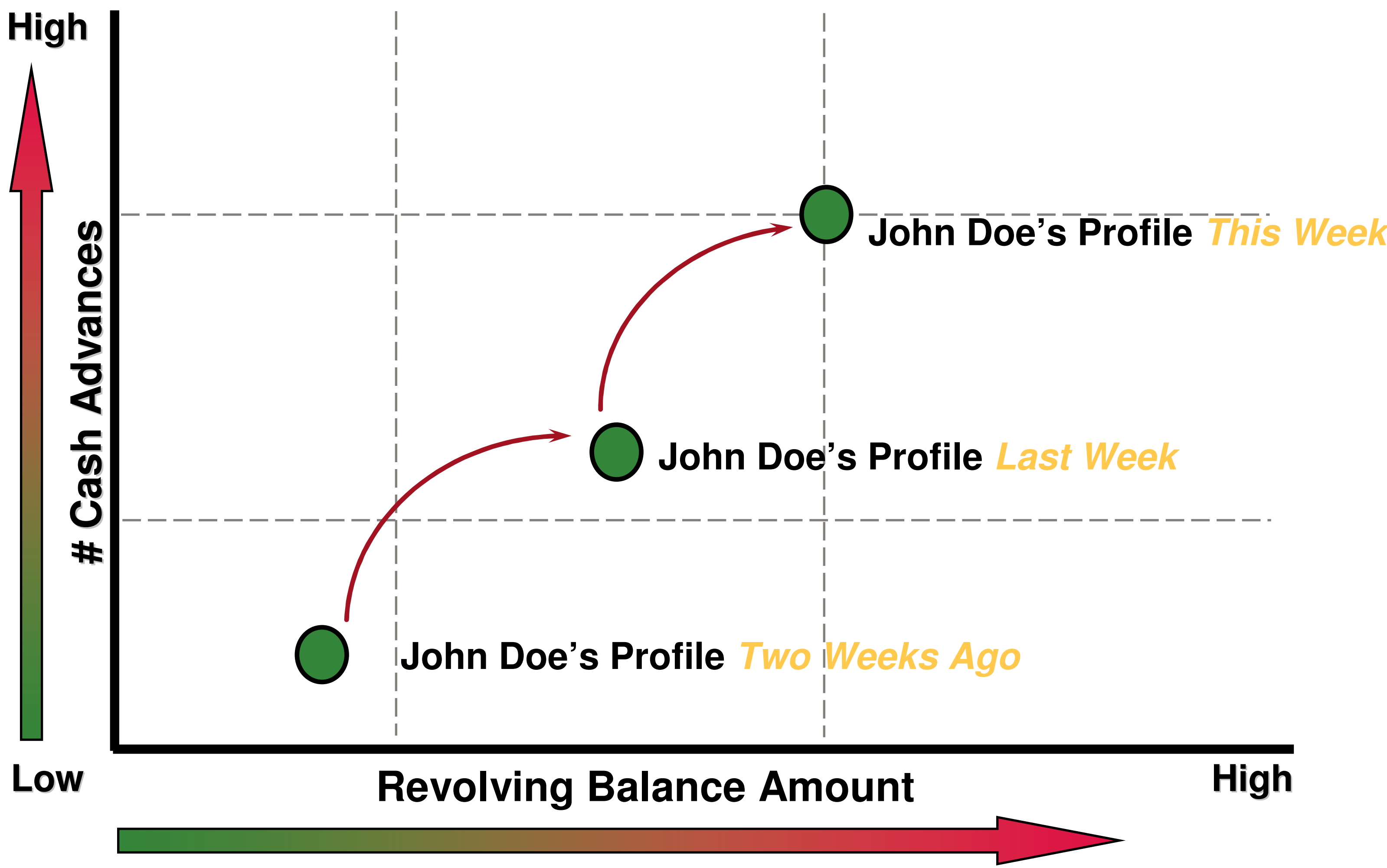


Slide 17

HF43 See email to Min asking for further explanation on this and next one

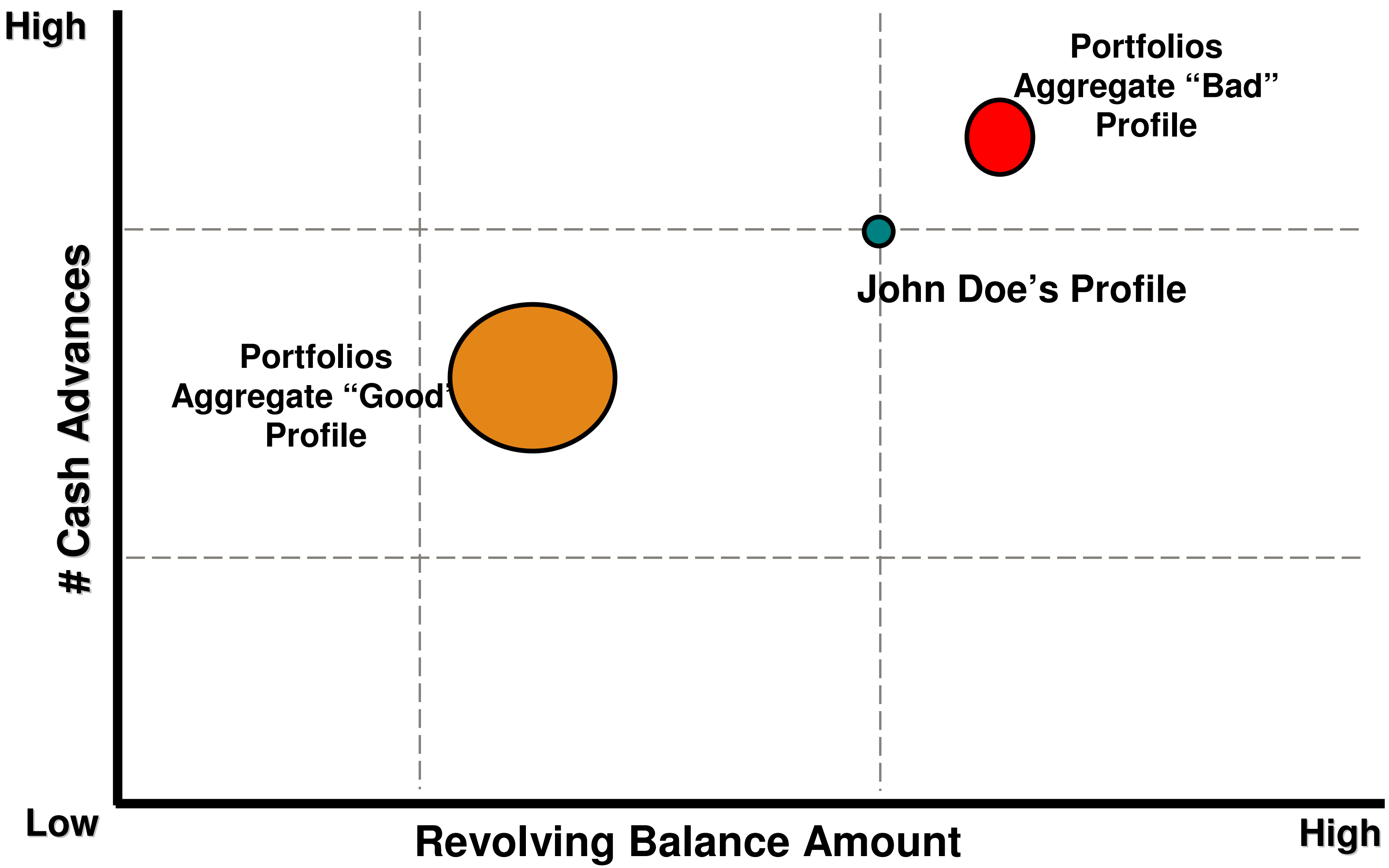
Hollis Fishelson-Holstine, 28/08/2005

Trending - Recognize Shifts in Individual Cardholder Behavior



HF37

Compare and Contrast Aggregated Behavior



Slide 19

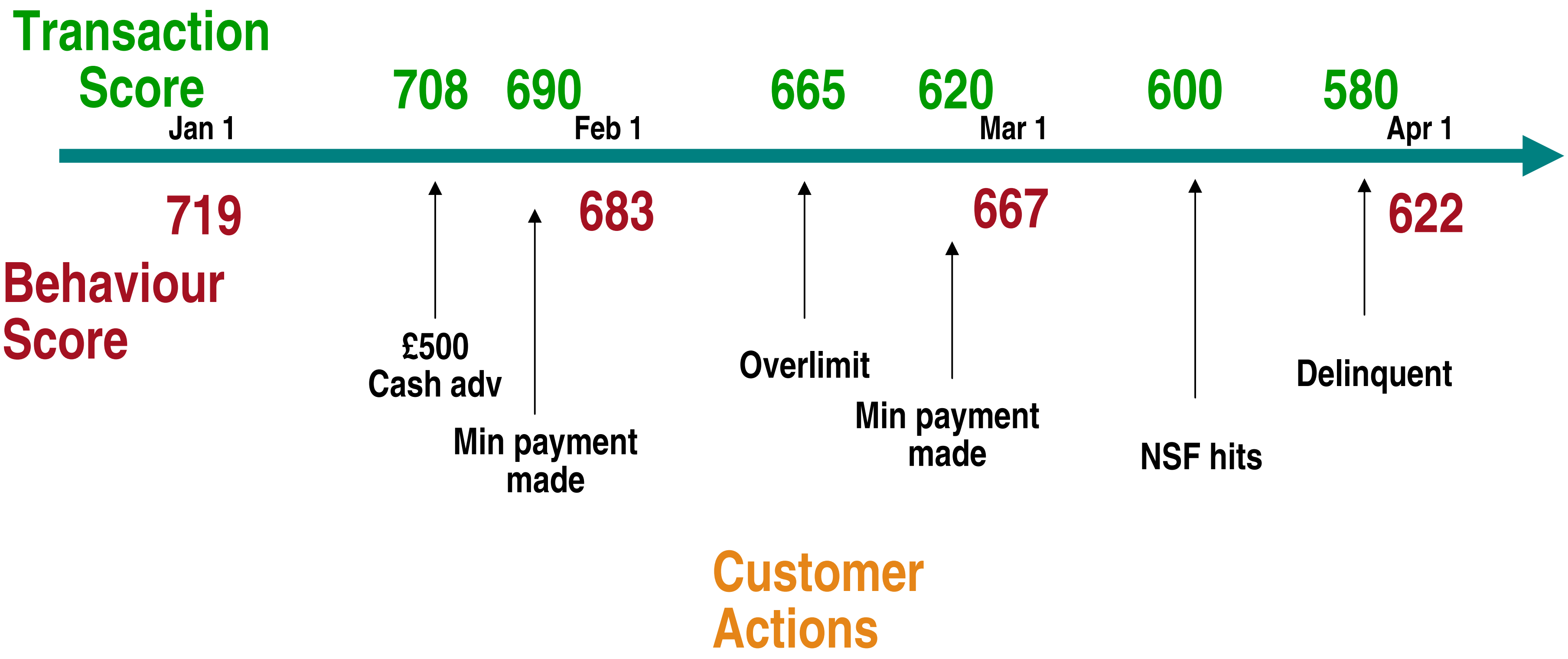
HF37 Andy will find TJ slide and add

Hollis Fishelson-Holstine, 24/08/2005

Agenda

- Motivation
- **Solutions**
  - Advanced technology
  - Increased understanding
  - **Better decisions**
- Challenges
- Case Studies
- Summary

Better decisions and more timely decisions



Triggers for Marketing Communication

	Home Improvements →	Home Equity Loan
	Car Broke Down →	Auto Loan
	Marriage →	Mortgage or Rental Insurance
	New Baby →	Life Insurance
	College Graduation →	Card Upgrade

Cross-sell solutions
Car loan trigger

Trigger definition

- Identification of large value motor repair transactions
- Identification of towing and crash repair transactions
- Identification of driving school and test fee transactions
- Trigger could include pre-approved decision criteria

Application

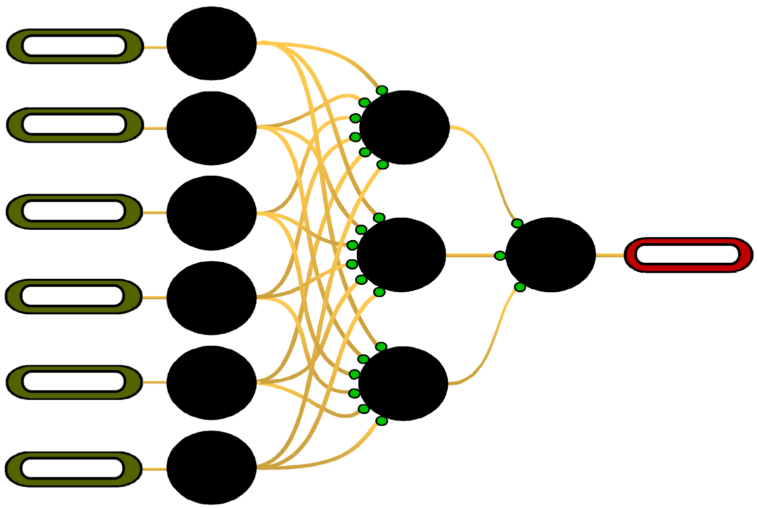
- Cross-sell for car loans

Agenda

- Motivation
- Solutions
- Challenges
  - Data volumes
  - Variable reduction
  - Stability
- Case Studies
- Summary

Cardholder Profiles

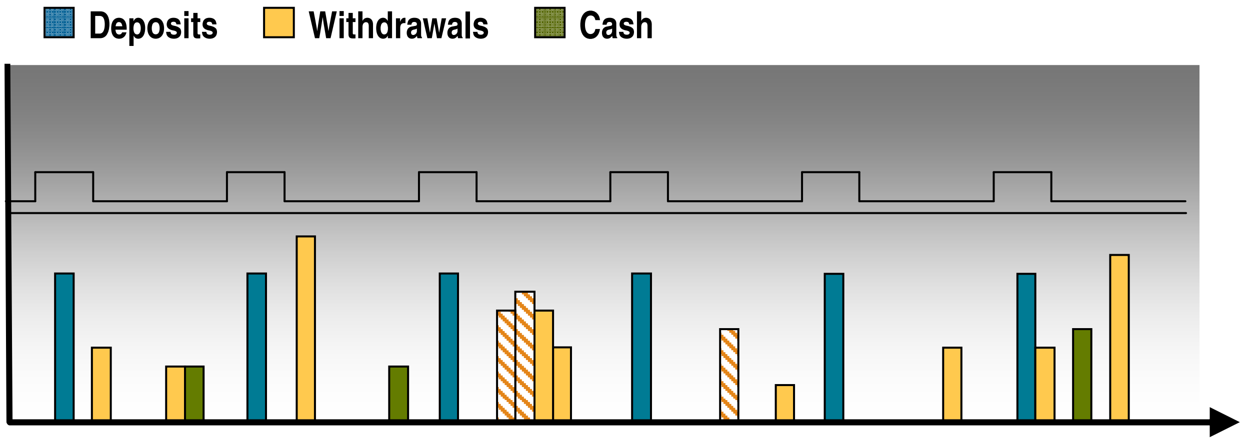
- Profiles utilize a wide array of information, examples
  - Long term typical historical behavior of the cardholder
  - Speed of spending
  - SIC/MCC codes frequented
  - Favorite shopping hours and days
  - Distance from home
  - How often the card is used
- Not just transaction data
  - Profiles also look at masterfile data

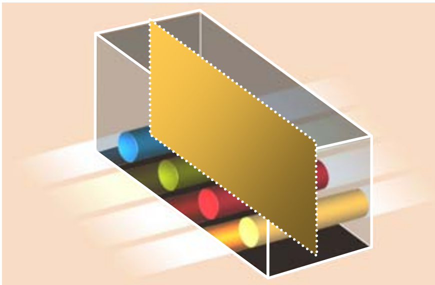


Analytic methods for periodic trigger design

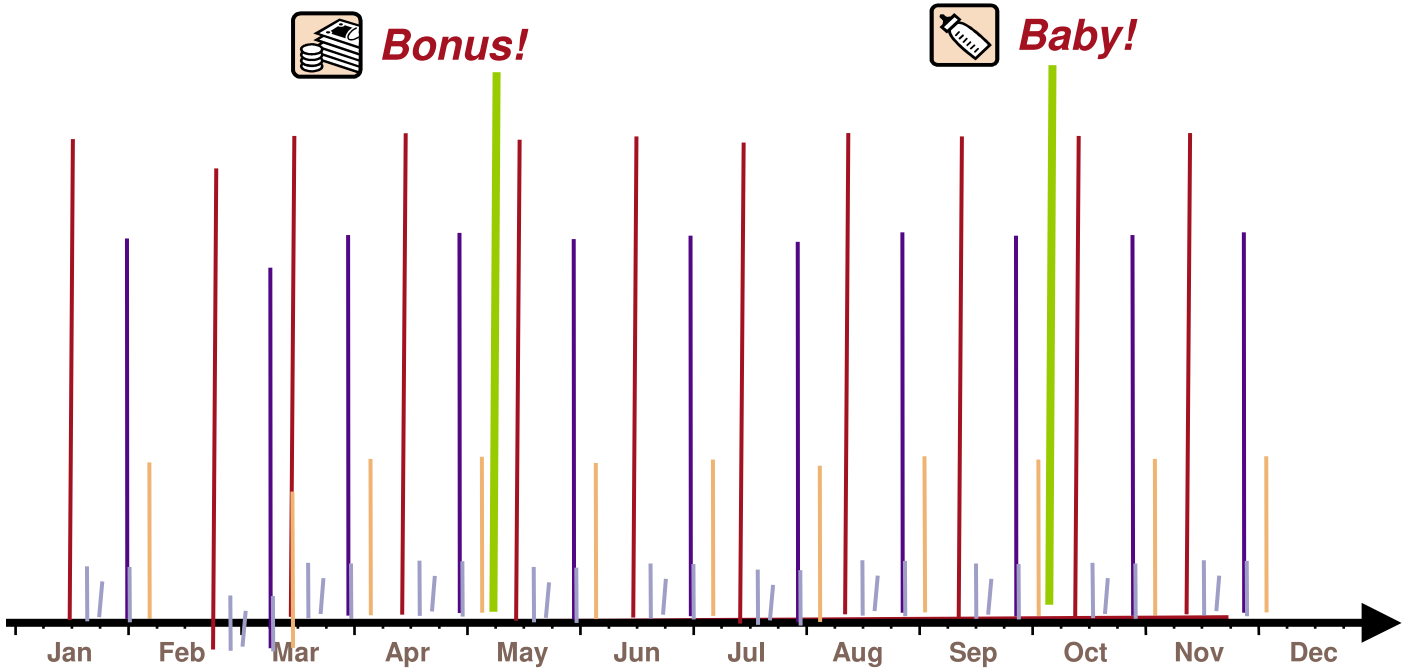
Comb or shaped filters

- Simple technique, easy to design
- Allows for the encoding of intuitive triggers or events
  - Facilitates encoding of payroll detection
- Can increase complexity with Fourier / Cosine transforms



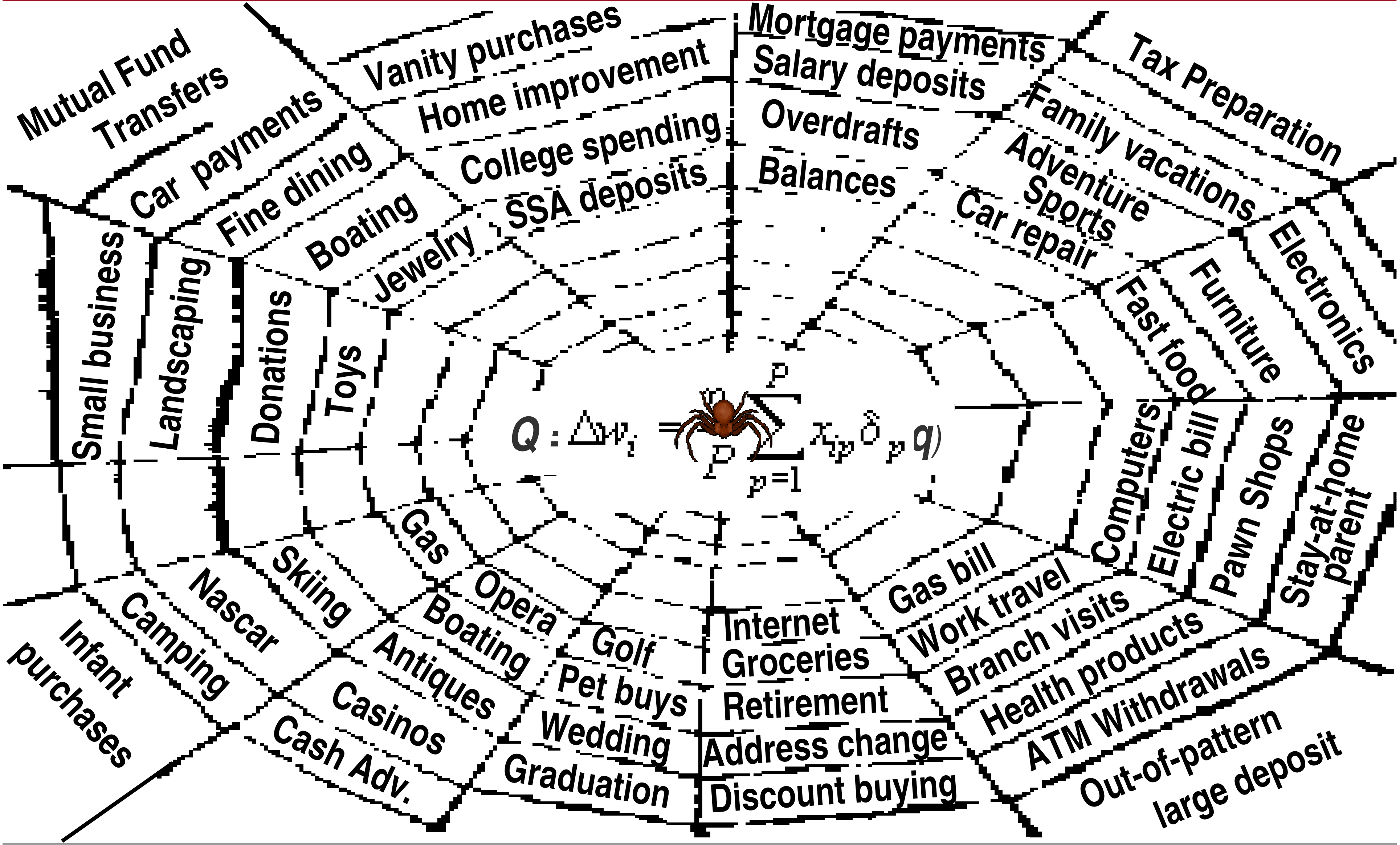


Pattern detection



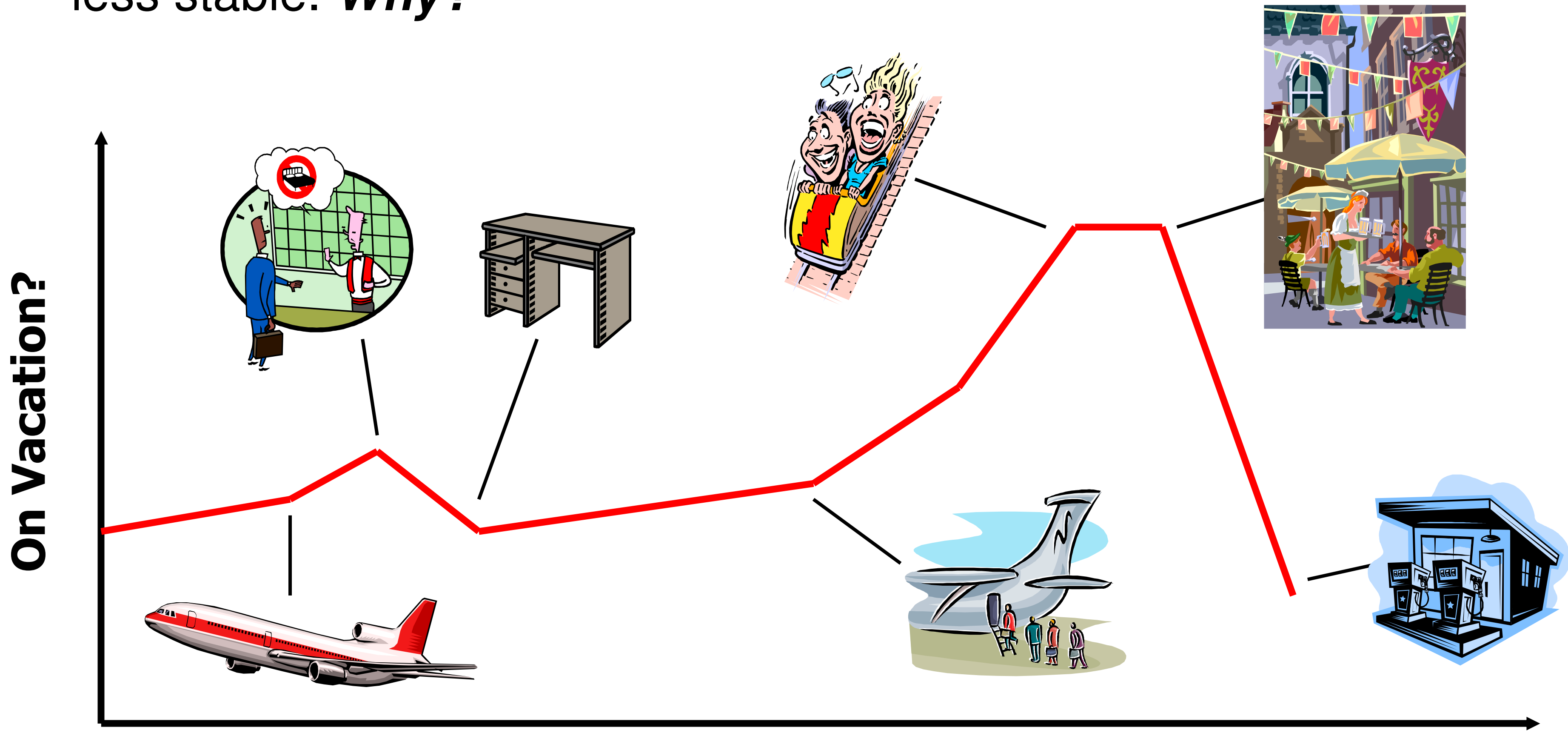
Salary	Biweekly fixed interval	Fixed amount, $2166.98
Mortgage	One month fixed interval	Fixed amount, $1245.32
Car Loan payment	One month fixed interval	Fixed amount, $399.00

Data Spiders find predictive variable combinations



Transaction Model Stability

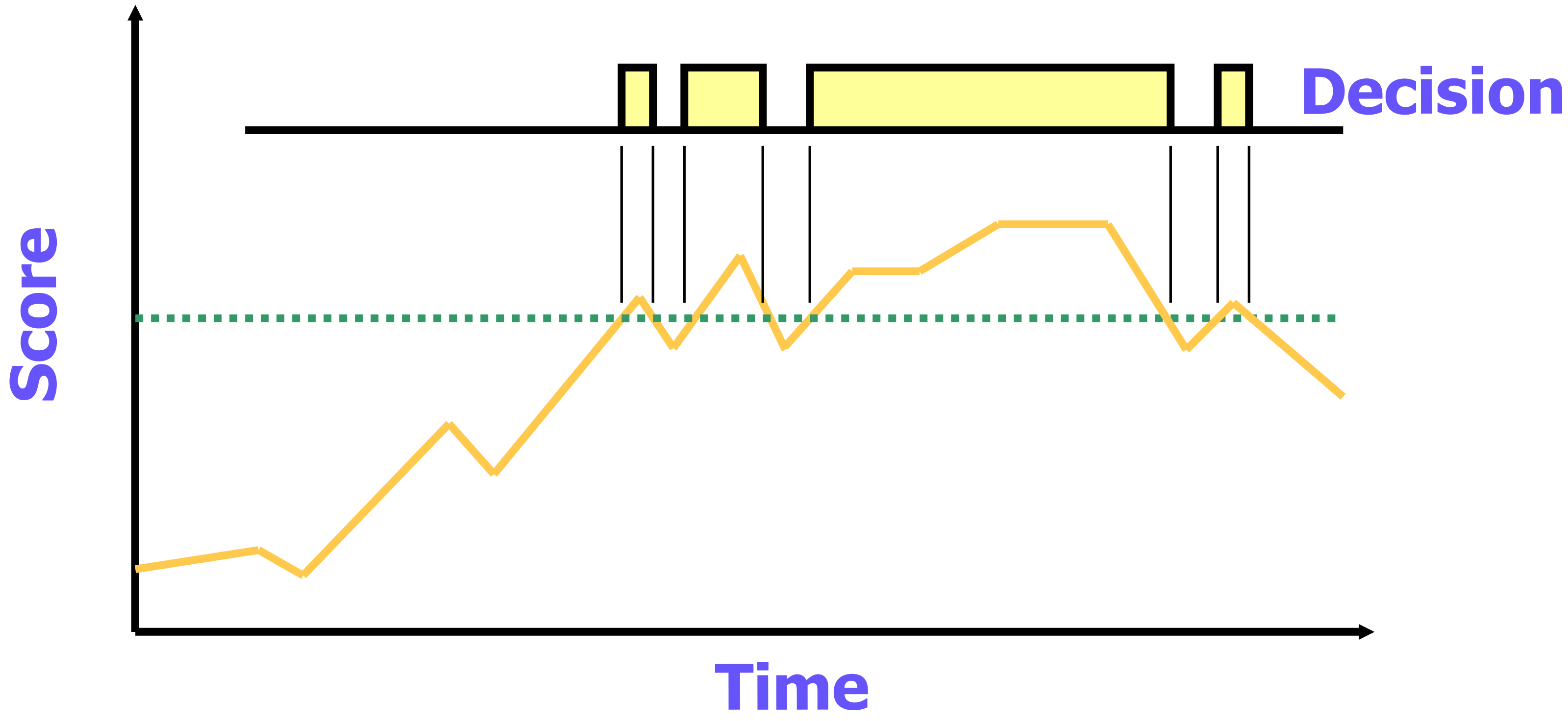
Transaction-based models can be very powerful, but they can be less stable. ***Why?***



HF26

Decision Hysteresis

Applying simple decision thresholds can result in unstable decision making



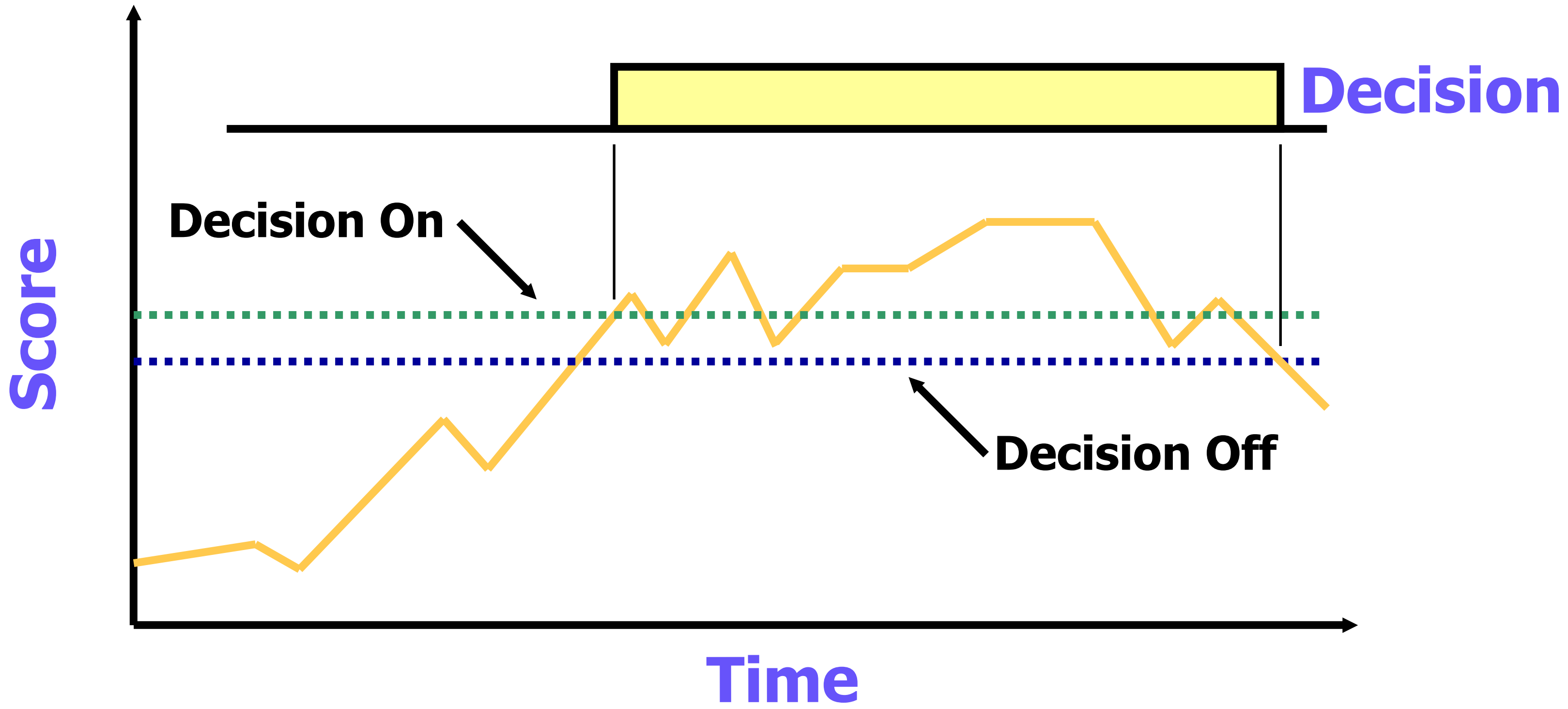
Slide 30

HF26 I thought you could use example in this and the next slide AND talk about value of profiles here

Hollis Fishelson-Holstine, 28/08/2005

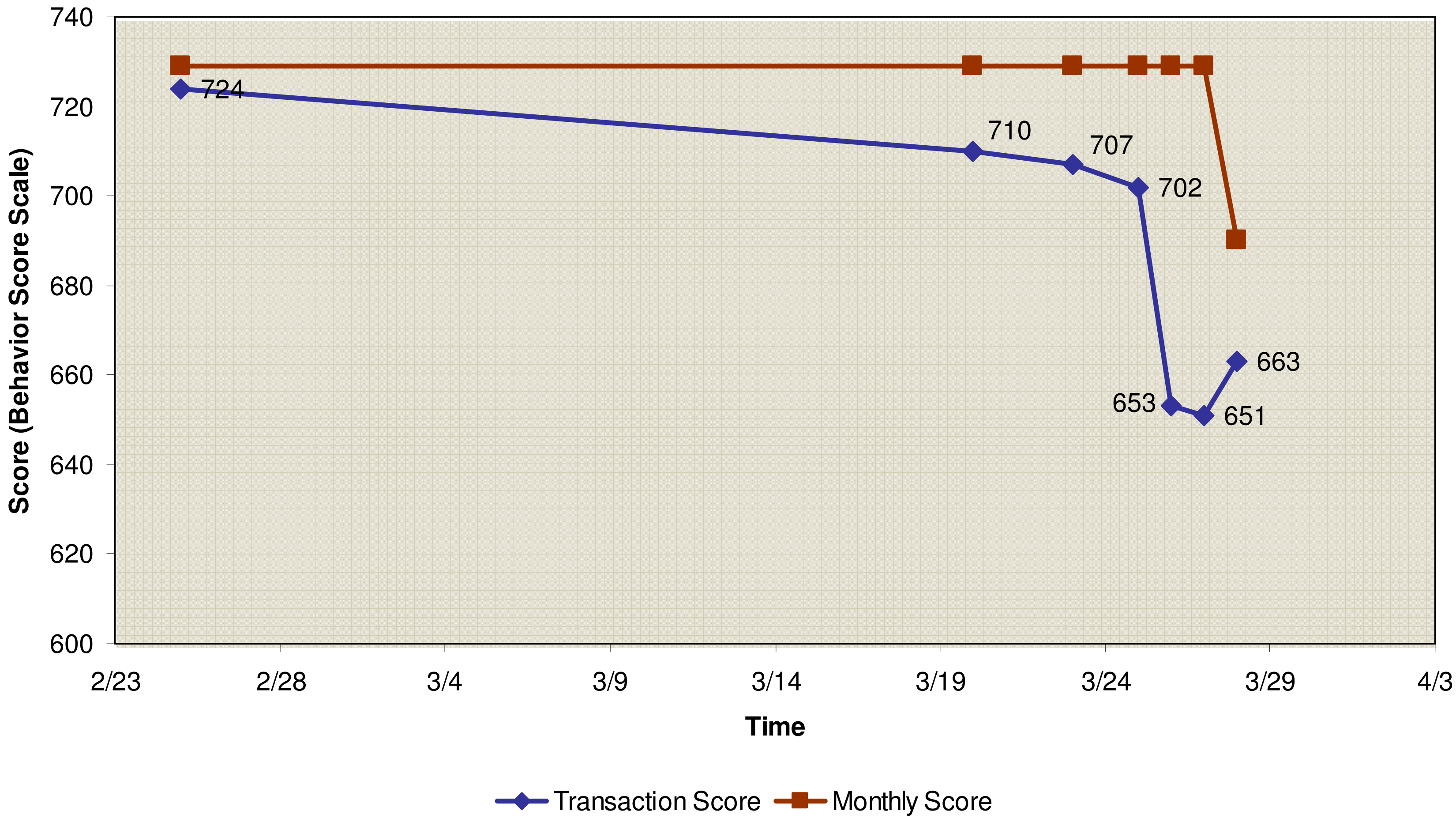
Applying Decision Hysteresis Solves This

Apply an upper and lower threshold, one to activate the decision, one to deactivate the decision



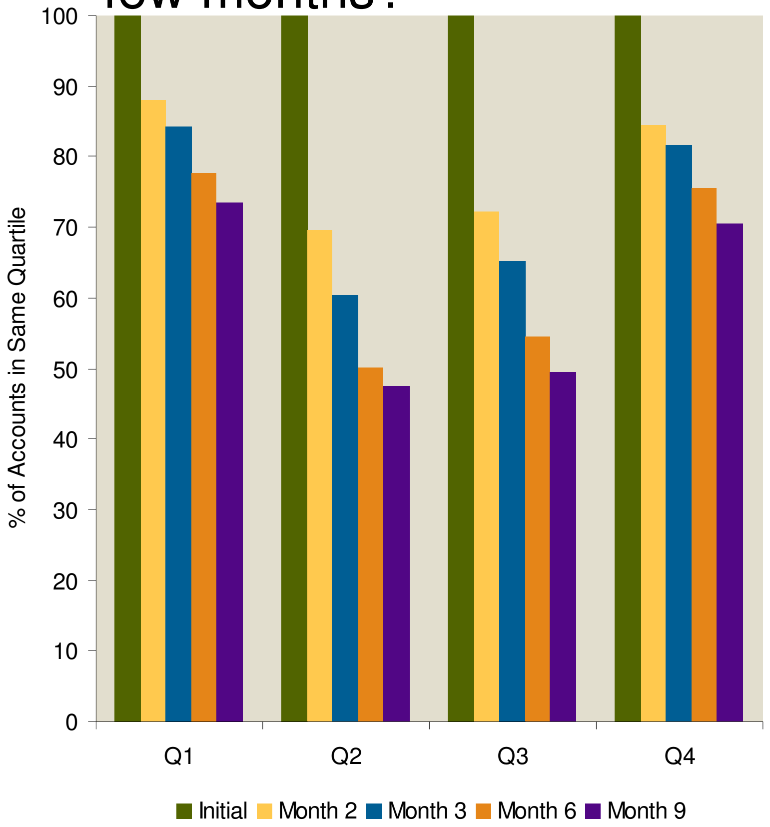
What about Stability?

- Real Data Example: How much do scores change during a month?

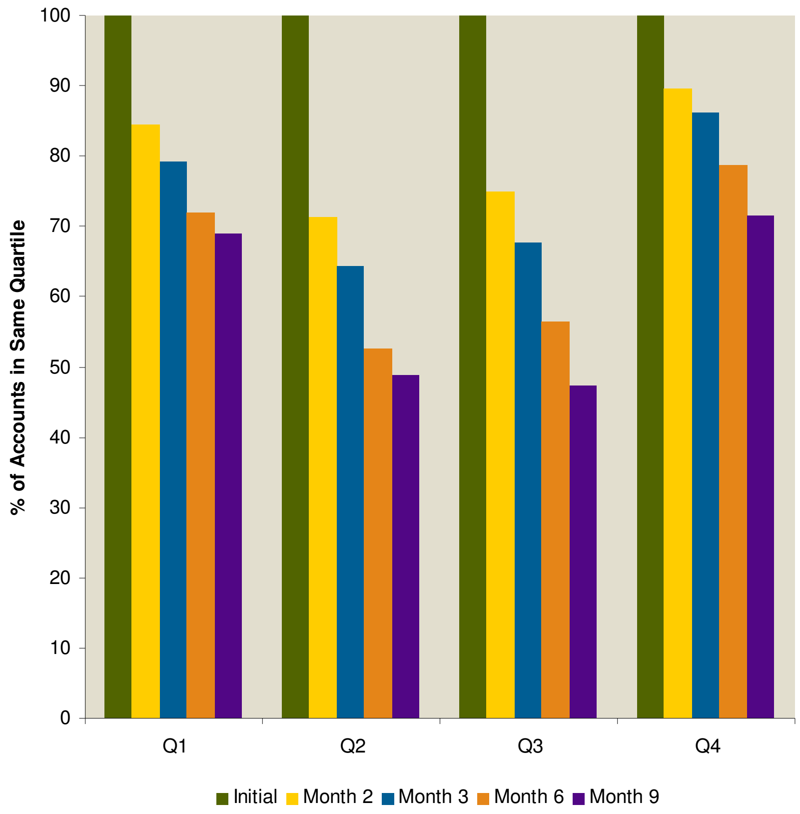


Long term stability – Transaction scores

- How much do TRIAD Transaction Scores change over the course of a few months?



- How much do Behaviour scores change over the course of a few months ?



Other Challenges

- Scalability
- Flexibility
- Deployment (models, triggers)
- Automation (test & learn)
- Cost

HF28

Agenda

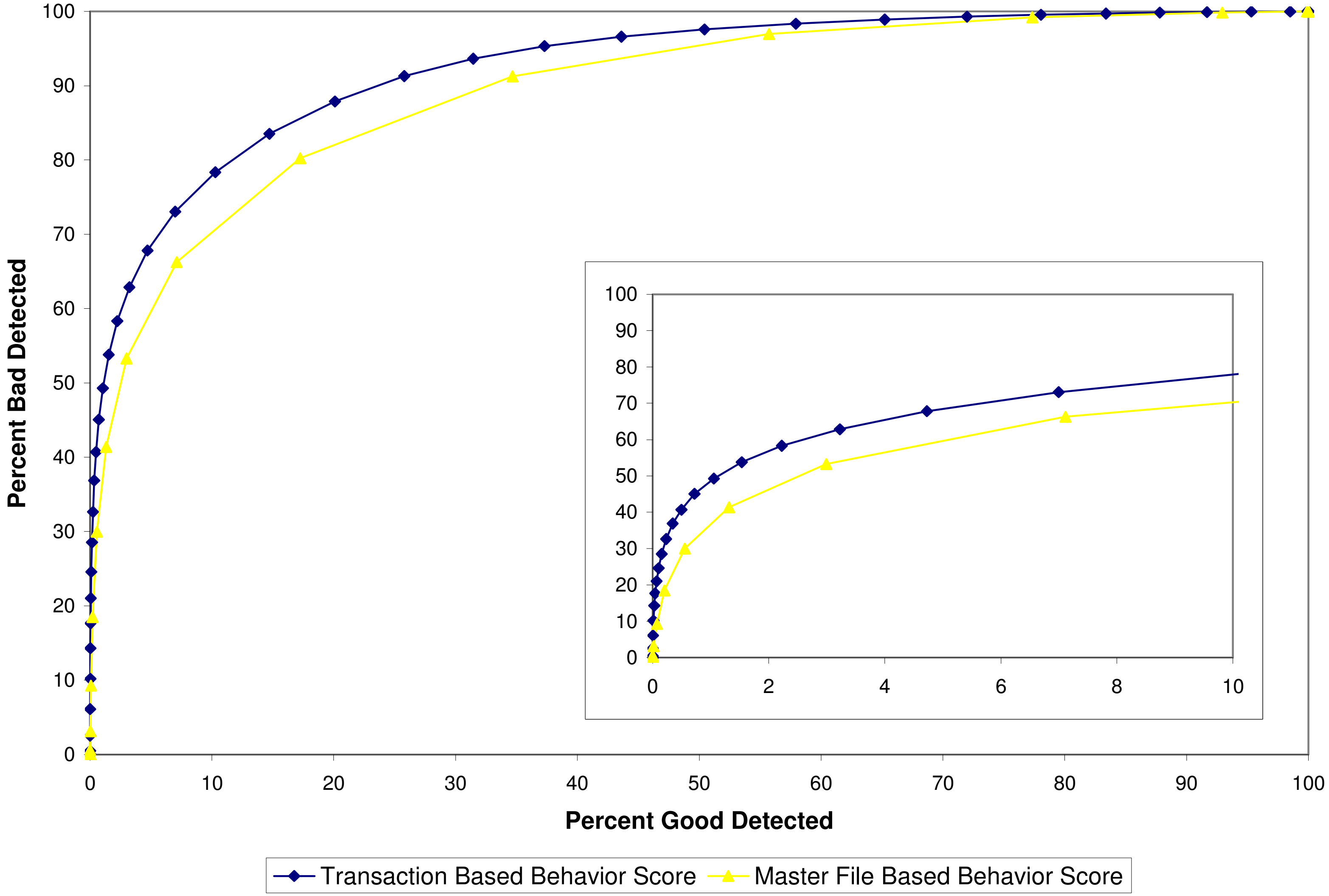
- Motivation
- Solutions
- Challenges
- **Case Studies**
- Summary

Slide 35

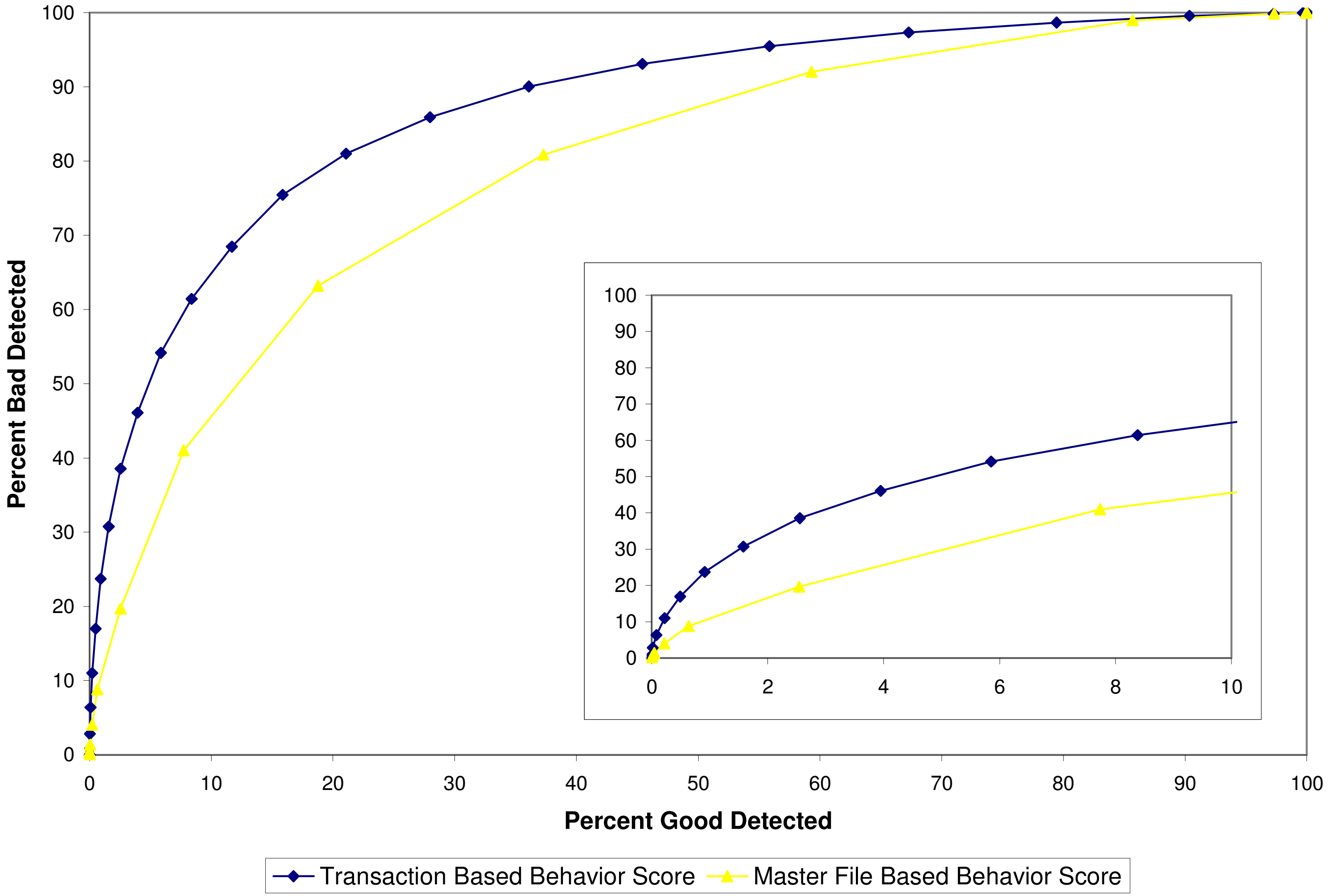
HF28 This is where I still really have work to do to condense, perhaps put in a more uniform format of "objective, challenges, results" (want that or different?)

Hollis Fishelson-Holstine, 24/08/2005

Credit Risk Model Performance Improvement



Young Account Segment Benefits More



What about Results?

- Real Data Example: Authorisations

Authorisations - In order Accounts		**Transaction**		**Monthly**
28%	***of overlimit authorisation attempts declined on current accounts***	**Score**		**Score**
# of current accounts				
# of attempted overlimit authorisations		188,052		188,052
% of transacting bads identified with score cut-off		67.6%		62.7%
% improvement in bads identified		**7.8%**		
# of bads identified with score cut-off		6,764		6,272
Additional bads identified		**492**		
Bads to charge-off ratio		65.0%		
Additional future charge-off identified by transaction score		**320**		
Average transaction amount		£170		
Balances saved from C/O		£54,390		
Monthly Authorisation Benefit for In order Accounts		£54,390		
Annual Authorisation Benefit for In order Accounts		**£652,676**		

What about Results?

▶ Real Data Example: Authorisations

Authorisations - One Cycle Accounts	Transaction		Monthly
27% ***of one-cycle accounts not eligible for authorsation***	**Score**		**Score**
# of one cycle accounts			
# of auth requests per one-cycle delinquent account	1.0		1.0
# of one cycle accounts requesting authorisation	107,500		107,500
% of total bads identified with score cut-off	63.7%		55.8%
% improvement in bads identified	**14.2%**		
# of bads identified with score cut-off	10,496		9,191
Additional bads identified with transaction score	**1,305**		
Bads to charge-off ratio	65%		
Additional future charge-off identified by transaction score	**849**		
# of future charge-off accounts requesting an overlimit authorization	849		
Average transaction amount	£170		
Total balances saved from charge-off	£144,247		
Monthly Authorisation Benefit for Underlimit One Cycle Accounts	£144,247		
Annual Authorisation Benefit for Underlimit One Cycle Accounts	**£1,730,958**		

First party fraud for large UK retail bank
The elements of first party fraud

- Definition:
  - When a customer takes credit with no intention of repayment

Behaviours

Open accounts with no intention of paying on them

Open accounts with false information

Boost credit limits – artificially and through manipulation of behaviour scores

False claims of fraud

Manifestation

First payment defaults or very early defaults

Excessively over-limit

First party fraud is big

First Party Fraud is significantly different than Third Party Fraud

First Party Fraud

- Perpetrated by customers
- Manifests as kiting, churning first pay default, skips
- No customer validation when it happens
- Very little control or detection of the problem

Estimated loss ratio = 100 basis points

Third Party Fraud

- Perpetrated by others
- Manifests as lost/stolen, counterfeit, forgery
- Customer confirms the occurrence of fraud
- Many controls and fraud detection systems in place

Estimated loss ratio = 10 to 12 basis points

A closer look revealed a common fraud scheme

First party fraud schemes:

(1) Account Opening

Individual opens account with some true and some false information or documents

(2) Transacts

Individual transacts heavily over time to give the appearance of creditworthiness

(3) New Limits Displayed

Individual qualifies for new overdraft limits and more loans

(4) Draw Facilities

Individual takes as much as they can from the current account and other loans

(5)Collections/Recovery

Account slips into collections, the individual in most cases skips or changes identity

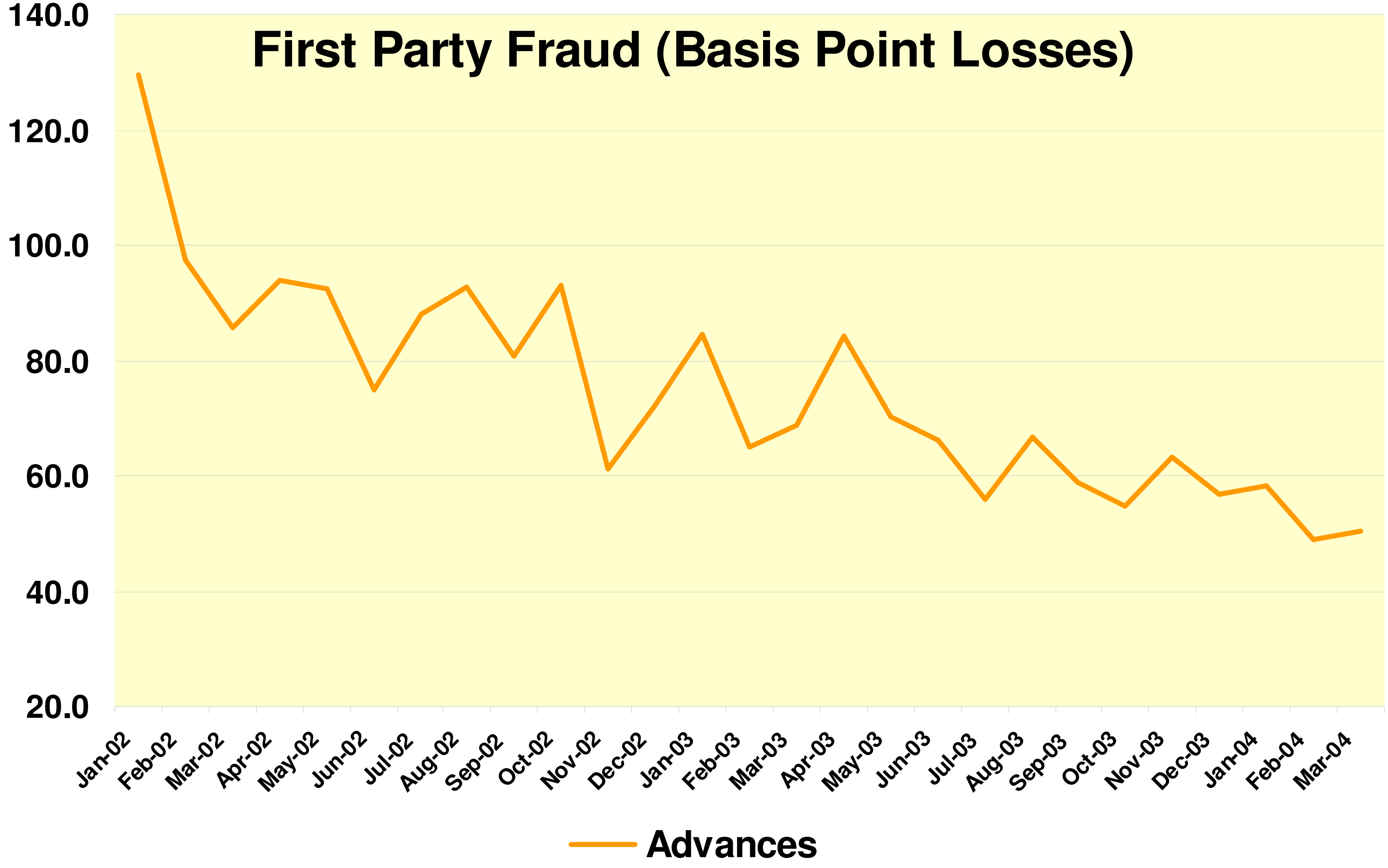
(6) Write-off

Account is written off, typically as a bad debt, not fraud since no known victim

Solution

- **Neural network models**
  - Use data from 48 sources (customer, account and transaction)
  - Score each account daily if a new transaction was made
  - Process millions of transactions a day
  - Utilise a sophisticated profiling engine
- **Case Manager**
  - Establish a holistic view into the customer relationship
  - Provide prioritisation of cases based on risk

We reduced First Party Fraud losses by over 50%

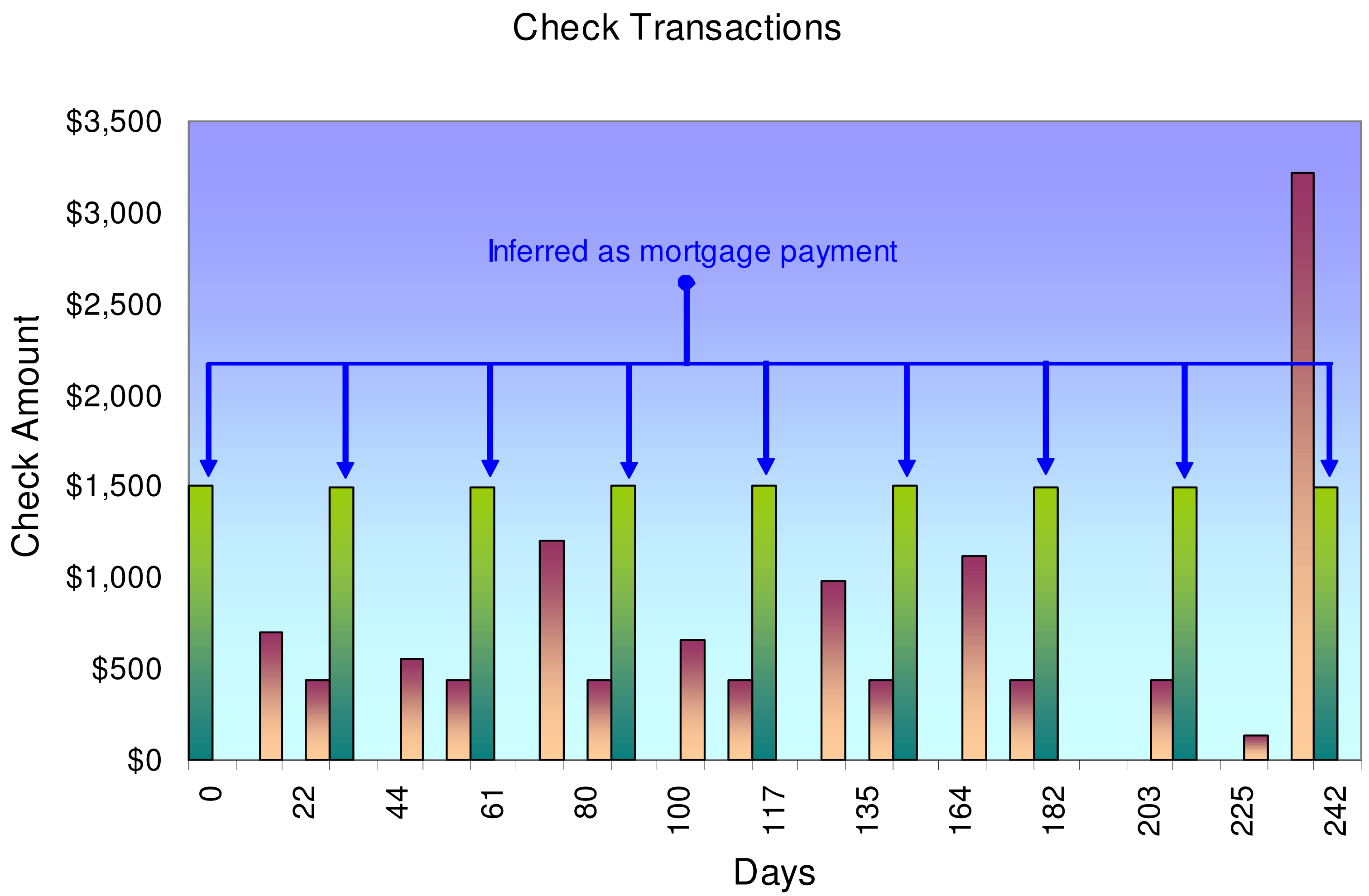


Cross-sell Home Equity Lines based on DDA (Current Account) Transactions and other data

- **Goal**: Derive Conversion Models, based upon customer transaction behavior, which can be applied to DDA customers for Home Equity Line of Credit Cross Sell Opportunities

- **Challenges:**
  - Identify existing mortgage holders
  - Timely deployment
  - Use Data Spiders to find predictive transaction characteristics

Illustrating the Mortgage Detection Algorithm



Examples of transactions found by Data Spiders

- Proportion of all transactions done via ATM over the last 210 days
- Time since the most recent home improvement transaction in the last 240 days
- Ratio of the number of txns at MCC 5814 (restaurants) in the last 20 days to the last 210 days
- Percentage of transactions greater than $1,000 in the last 240 days
- # of checks greater than $50 in the last 240 days
- # of days since the most recent home improvement transaction in the last 180 days
- Maximum amount of all home improvement transactions in the last 240 days

Results

- Typical Home Equity Line worth \$400 - \$700
- Bank with 3 million checking accounts
- Increase by 2% the checking account customers with Home Equity LOC
- Increase profit \$24 million

Context Vector Technology Applied to Collections Modeling

- Once in collections beyond 1 cycle delinquent little structured transactional data
- Difficult to predict customer behavior in collections cycle
- Little or no differentiation in customer treatment
- Late stage collections tactics not well used by issuer

Mathematical representations of meaning

Learning meaning from text data through analytics

"You shall know a word by the company it keeps"

--(Firth,1957)

Context defines meaning in a robust way

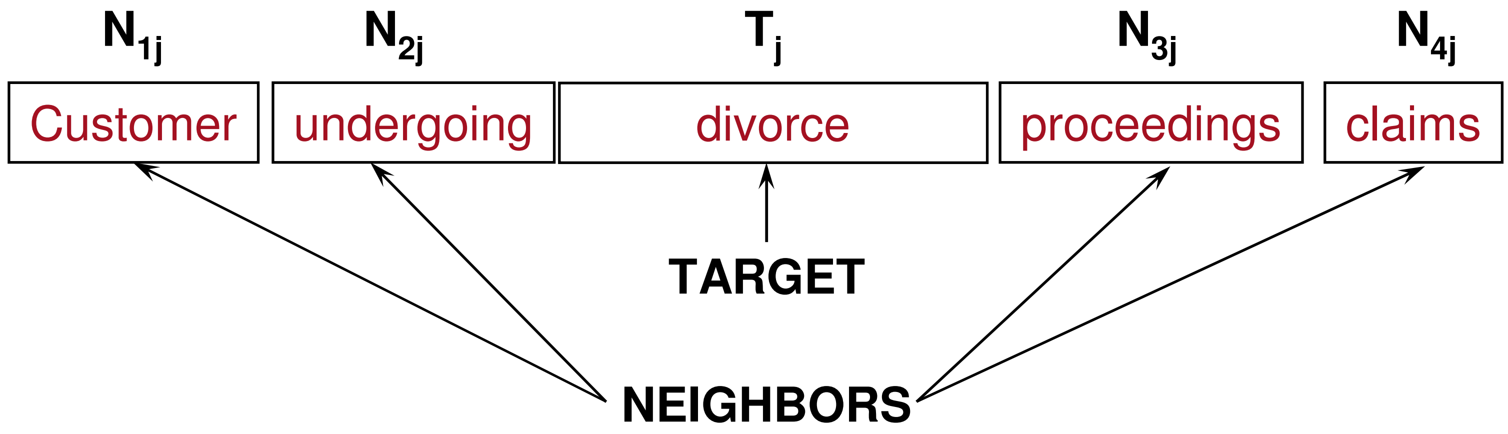
- What is this word?
  - *iprmoetnt*
- Now guess what this word is:
  - "the only *iprmoetnt* thing is that..."
- Now read this entire sentence:
  - *'Aoccdrnig to a rscheearch at an Elingsh uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht frist and lsat ltteer is at the rghit pclae and the cotenxt. The rset can be a toatl mses and you can sitll raed it wouthit porbelm. Tihs is bcuseae we do not raed ervey lteter by itslef but the wrod as a wlohe. ceehiro.'*

Meaning = Statistical Context

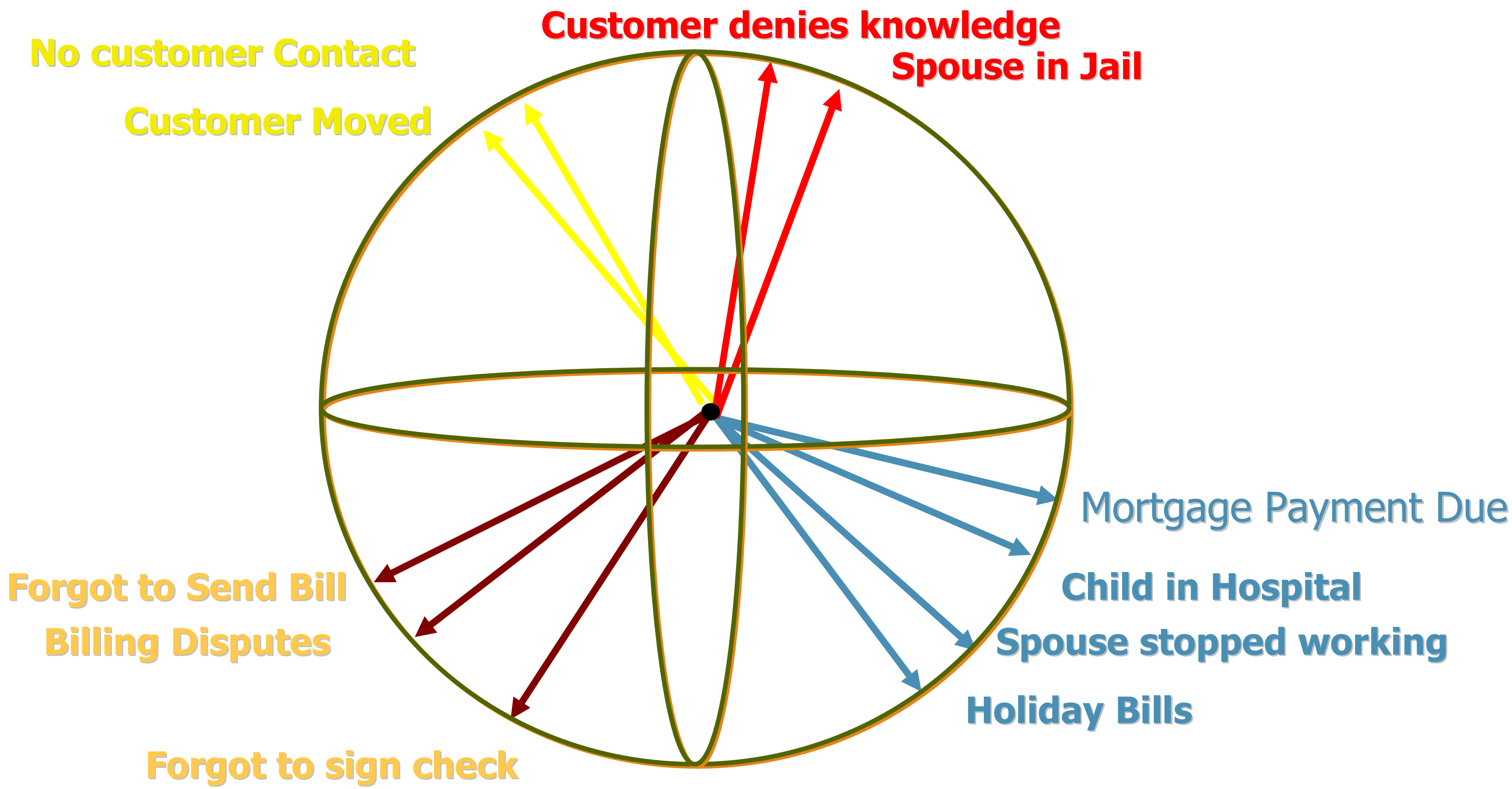
Source Text

Customer undergoing divorce proceedings. Claims spouse is responsible.

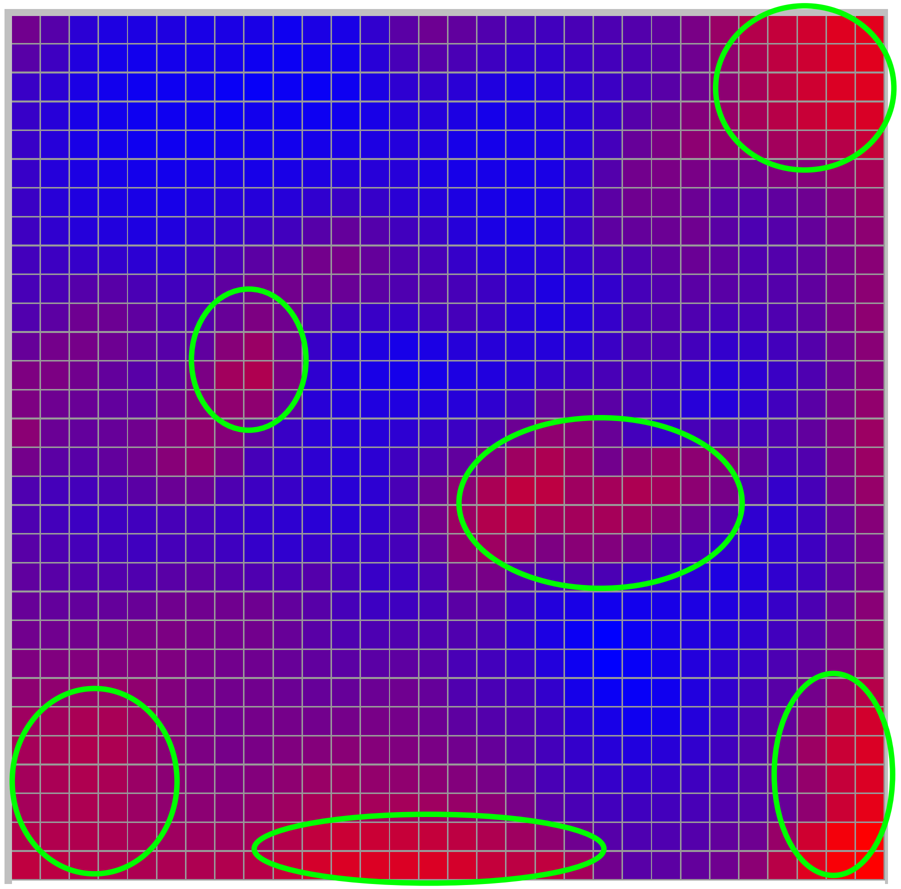
Five Stem Window



Collections Notes Vectors ➡ Cluster into segments with different payment potential



Concept Space for Collection Notes
Why do borrowers don't pay?
Text → "Reason Codes" → Appropriate Actions



HF45

Context Vector Technology Applied to Collections Modeling

▶ Portfolio Receivables Valuation pilot

- Medium sized bank
- 590K accounts with text notes (58% of all accounts)
- 1 to 280 comments per account with average=14
- Target:
  - 12 mo collection (as % of owed) on date of each comment
  - First 3 months of comments used for modeling records

Slide 55

HF45 Re-write - incorporate other slides on set-up

Hollis Fishelson-Holstine, 24/08/2005

Context Vector Technology Applied to collections Modeling
Dramatically Better Receivables Valuation Models

Better lift:

Model KS with no-text =27.8

Model KS with text =52.5

Increase in model accuracy driven by text =88%

More accurate prediction leads to lower reserves:

Total actual dollars = $33,469K

Predicted with text = $42,021K (125% of actual)

Predicted with no-text = $67,460K (201% of actual)

HF46 Why are top and bottom backwards? - thought weren't this good
Hollis Fishelson-Holstine, 24/08/2005

Summary

- Increasing wallet share while managing risk and fraud requires new techniques for better decisions
- Decisions can be improved through better technology, more information, and more timely information
- Technology from a variety of sources can be used for commercial benefit
- Transaction data
  - captures important events and identifies changing patterns of behavior
  - is further enhanced through text or unstructured data
  - allows for more timely action
- Working with transaction data is challenging, BUT
- The results are large!