Building, refining and scoring financial transaction sequence analytics

Robin Way, Corios LLC, Portland, Oregon

ABSTRACT

At the A2010 and A2011 SAS analytics conferences, the author introduced the audience to the practice of transaction sequence models, applied to marketing cross-sell, risk management and account servicing case studies in the financial services industry. Sequence model discovery yields important insights about what financial transaction-based events happen together, in which order, how often. The response was very positive, yet the audience wanted more: show me how to develop and refine the data for these models; show me how to actually build these models using tools of the trade; and show me how to apply the results of these sequence models as predictors of some meaningful customer outcome, such as response to a risk management or marketing treatment.

In this presentation, the author will share the key requirements for practical, in-field refinement and scoring of sequence transaction models. Refinement of transactional records by in-database filtering, cluster sampling, and time-series transformation techniques will provide insights into real-world practices to get transaction data ready for sequence model discovery. Next, development of sequence model candidates and deployment of sequence scoring will provide the audience with the identification of which customers exhibited specific sequence patterns. Finally, incorporation of the sequence scores into a subsequent cross-sectional predictive model will yield the measurable payoff: using sequence scores as leading indicators of response to a treatment, conditioned for transaction-level behavior in the period leading up to the treatment.

INTRODUCTION

A sequence (in the context of this paper) is a series of events. As human observers, we observe event sequences everywhere, all the time. Our brains are well tuned to identifying sequences that are common to us, and in noting events that vary from the expected pattern. In the financial services industry, there is tremendous potential for analysts to enhance the way that transaction-based event sequences are discovered, recorded, and used as potential predictors of important events like response, delinquency, default, and financial crime. This paper introduces the concept of event sequences characterized by financial services transactions based on real-world experience, and the augmentation of analytic techniques through the use of sequence analysis and scoring.

WHAT ARE SEQUENCES?

Let's start by identifying a transaction pattern familiar to those steeped in the data mining industry lore, which is not a sequence, but an association, more commonly called a “market basket”. This is the notion of beer and diapers. The origin of the beer and diapers metaphor is not important here but makes for fun reading on the Internet. By studying retail transaction patterns of items placed in a market basket, we can identify customers who have placed pairs of retail items in their market basket. If the order in which they placed those items in the basket, or the order in which they purchased them, is not important to the analyst, then we call the combination of beer plus diapers an “association rule”.

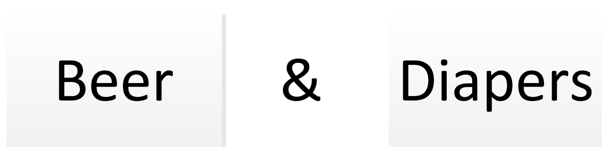


Figure 1: An association or “market basket” rule

In contrast, a sequence is a series of items placed in the basket or purchased by the customer in a specified order; the order in which this sequence occurs is important to the analyst. For instance, first the customer places beer in the basket, then time passes, then the customer places diapers in the basket. The passage of time, and the retention of information about which events occur first, second and so on is what differentiates a sequence rule from an association rule.

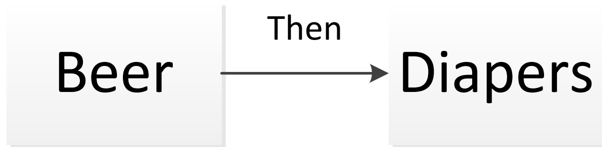


Figure 2: A sequence rule

SEQUENCE CHARACTERISTICS

In a generalized sense, one can characterize sequences as being composed of a **precedent** (what happened first) and an **antecedent** (what happened next). In practice, a precedent event or antecedent event could be an individual transaction, a series of commonly occurring set of transactions, or a generalized behavior state exhibited by the customer. On their own, combinations of precedent and antecedent rules are descriptive, but not inherently predictive. For that reason, I think it makes a lot of sense to place sequence rules in a business context that aligns the useful descriptive information with a customer strategy of interest to the business. This gives rise to a third sequence concept that I call the **consequent** (what should I do about it).

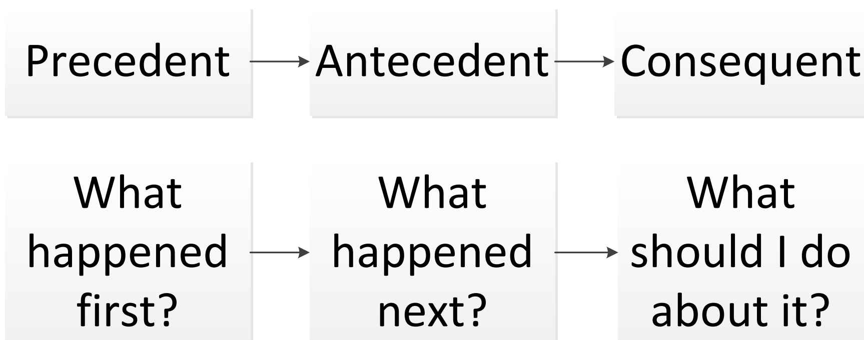


Figure 3: Sequence concepts: precedent, antecedent, consequent

There have been many excellent research initiatives performed in the area of sequence analysis, though most of them have been focused either on genomic research or website traffic patterns, or on the computational approaches required to discover interesting sequences. There have not been as many research initiatives focused on the context of sequences in a business domain like financial services, and fewer yet that focus the interestingness of potential sequences as predictors of customer behavior in response to a financial service firm customer strategy. This is why the concept of the consequent component of sequence analysis is so important; it offers the potential to transform a descriptive analysis into a predictive one, and thereby generate direct business value to the firm.

In an abstract sense, event sequences can be found nearly everywhere in the natural world, and this is true also for the financial services industry. The human mind is a keen detector of event sequences, and particularly so when real-world phenomena occasionally exhibit qualities that vary from what we expect to observe. Of course, when presented with a database consisting of millions of transactions, and the tens of millions of prospective event sequences comprised by those transactions, we need to rely on computational engines to help us identify, record and discover those sequences of the greatest interest.

In a practical sense, event sequences help tell the narrative story of a customer's experience with the financial services firm. It is customary in most functional business processes in a bank, insurance, payments or brokerage setting to aggregate customer behavior of various types to the calendar month or billing statement level. I suspect this is a matter of tradition: cost of storage, processing time, and the dearth of transaction-oriented computational algorithms in years past naturally limited the exploitation of transaction data below the level of the customer-month.

As those barriers are knocked down by lower cost, faster and more powerful computational techniques, the possibility and practicality of event sequence analysis techniques should naturally rise; and yet, there remains a fundamental lack of awareness of event sequence capabilities in day-to-day predictive analytics practice.

In the financial services industry, analysts have become accustomed to the conventional customer-month data warehouse record, yet the real customer story transcends this artificial boundary. Consumer behavior is driven by factors more far-reaching than just household liquidity and the ability to borrow. Our capacity to characterize consumer behavior in its naturally occurring order and cyclicality has been tremendously increased by the advent of sequence analytics, if only we knew how to use them.

CULTURAL EXAMPLES OF SEQUENCES

We observe sequences all the time in our daily lives, often without consciously recognizing them. Some event sequences are very short-term instances, lasting only a few seconds, such as the steps required to pay for a cup of coffee at a café: open wallet, hand debit card to cashier, politely decline a paper receipt, retrieve debit card, receive the requested beverage. Other event sequences occur over long extents of time: consider the common stages in a person's life cycle, composed of birth, school, career, marriage, family, retirement and death. Some individuals experience these life stages in a different order, but we'd all accept that the sequences are common enough to make interesting comparisons about the relationship of lifecycle stages, and their sequences, with regard to the individual's life experience.

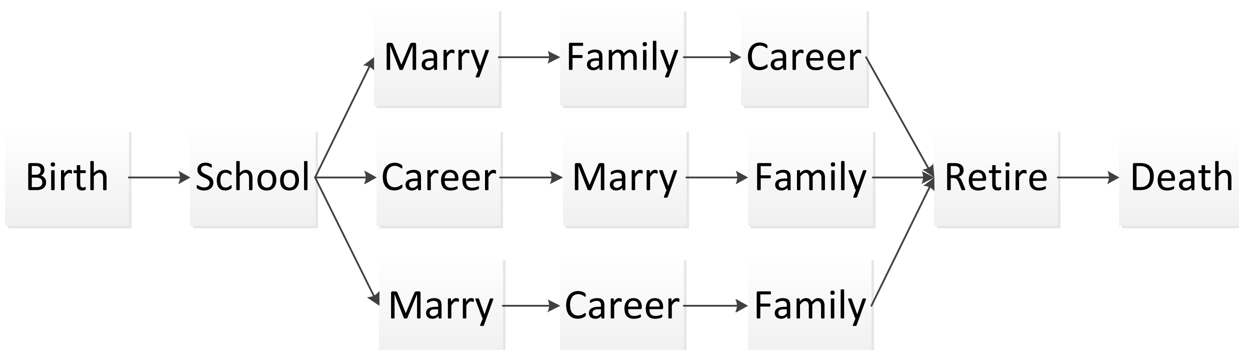


Figure 4: Life cycle stages as sequences

Relax your assumptions about our subject matter in the financial services industry for a moment and think about all the places you can observe sequences. Every morning, you probably follow a common habit in waking, starting your day, doing familiar things and visiting familiar places. The author happens to be a fan of the Beatles, who wrote a song called “Day in the Life” (from the album “Sergeant Pepper’s Lonely Hearts Club Band”), that captures this idea nicely. The lyrics of the song in the first stanza depict the protagonist’s life on a regular day.

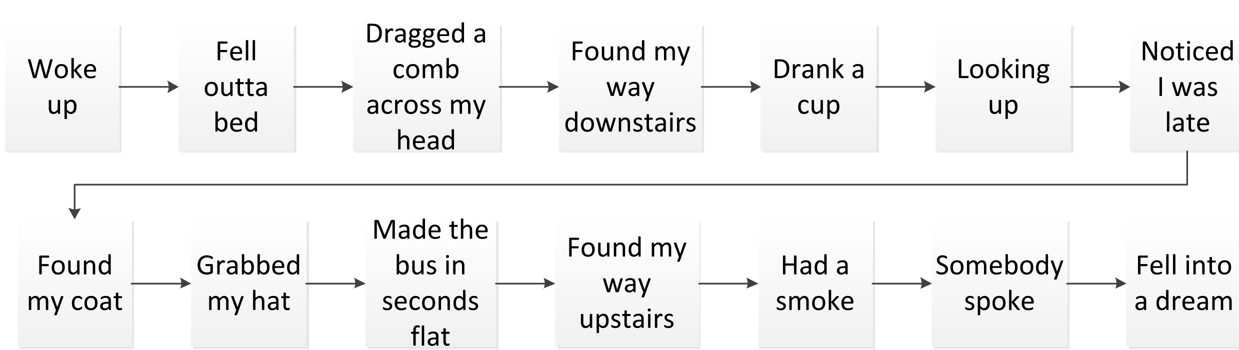


Figure 5: Song lyrics as sequences (Lyrics: John Lennon & Paul McCartney; © Sony Beatles Ltd.)

If you really let your mind wander, you can see how that single day could be extended into a series of days with a similar set of events. Likewise, any observed sequence could be only one event in a chain of a larger sequence.

And at a more discrete level, think about the Beatles' song itself and its structure; a song's lyrics are composed of stanzas with common length, and each stanza composed of lines of common length and meter. A song's music is composed of sections, which are composed of measures, which are composed of beats. These are all commonplace examples of sequences, which are so familiar and comfortable to us that we take them for granted. The human mind is exceptionally efficient at observing and categorizing events and sequences. In particular, our minds are well honed to detect the variations from the familiar sequences that characterize the events in our daily lives, work and culture.

FINANCIAL SERVICES EXAMPLES OF SEQUENCES

Let's turn to some more concrete examples of event sequences of importance to financial services firms, and in particular we'll use debit and credit payments as our context. Anyone who's worked in the payments industry will recognize the event sequence patterns represented in the following table. Consumers use debit and credit on an increasing scale to buy goods and services, and payments processors and issuers maintain vast data warehouses containing transactional histories at the cardholder level. Each transaction record contains the cardholder ID, merchant ID, transaction date, transaction volume (e.g., dollars and cents), and additional information specific to the merchant and payment device. By observing common sequences of these transaction types, we can quickly develop an extensive list of purchase patterns. These patterns can be differentiated by a number of attributes: frequency, periodicity, time span, dollar volume, the predictability of event sequence onset, and the conditional predictability of the antecedent event.

Event and sequence types	Examples	Frequency	Periodicity	Timespan	Dollar volume	Onset predictability	Antecedent predictability
Large monthly payments	Mortgage, rent, auto, phone, cable, insurance, installment loans	♦♦	Monthly	◷	$$$	✓✓	✓✓
Small monthly payments	Newspaper, magazines, transit, specialty subscriptions, Netflix	❖	Monthly	◷	$	✓✓	✓✓
Varying-size monthly payments	Credit payments, balance transfers, revolving loans	♦♦	Monthly	◷	$$	✓	✓
High-velocity financial transactions	Credit purchase, debit purchase, ATM cash withdrawal, online banking statement, brokerage position inquiry	❖	M/W/D	◷	$$	✓✓	✓✓
Weekly staples	Groceries, gasoline, parking	❖	Weekly	◷	$	✓✓	✓✓
Daily sundries	Coffee, breakfast, lunch, parking, transit	❖	Daily	◷	$	?	✓✓
High frequency small ticket items	Coffee, newspaper, muffin, beer, parking, pizza, take-out food	❖	Daily	◷	$	?	✓
A night out	Cocktails, dinner, theater tickets, dessert, cab fare	♦♦	M/SM/W	◷	$$	??	✓✓
High-expense hobbies	Car enthusiasts, fitness, sporting events, home theater, gardening, DIY home improvement	♦	M/SM/W	◷◷	$$$	??	✓✓
Short-span specialty retail & electronics	Store or catalog clothing, books, PayPal/Ebay transactions, iPhone apps, Kindle downloads	♦♦	M/SM/W	◷	$$	?	?
Business travel	Airfare, hotel, rental car, meals, parking, cab fare, internet access	♦	M/SM/W	◷◷	$$$	??	✓✓
Family activities	Movies, sports, shopping, camping	♦	Seasonal	◷◷	$$	?	✓
Three-day weekends and seasonal holidays	Gasoline, take-out food, motels, amusement parks, movie theaters	♦	Seasonal	◷◷	$$	??	✓
Infrequent long-span events	Family vacation, home remodel, relocation	♦	???	◷◷◷	$$$	???	✓
Infrequent short-span high-dollar	Furniture shopping, car repair	♦	???	◷	$$$	???	✓

Figure 6: Financial transaction event sequences associated with consumer spending patterns

Imagine that you are employed as a product manager for a credit issuer focused on targeted incentives for consumer credit purchases tied to vacation travel. The vacation patterns exhibited by cardholders with varying lifestyle choices, such as presence of children, disposable income, prior history of domestic versus international destinations, and timing of past vacations would all be useful indicators for a segmentation strategy of these cardholders. The most powerful information for identifying the ideal timing of the targeted incentive could be based on knowing the leading indicators of a pending vacation travel investment decision. You'd want to identify individuals making these decisions early enough to influence the choice of air carrier, lodging, rental car, restaurant, event and amenity selections in advance of the individual already having committed to those choices. Event sequences based on leading indicator purchases, and the timing of the onset of these purchases, form a critically important basis for when to target an incentive for those segments of cardholders.

The good news is that the development of these event sequences and their leading indicators is possible and practical. What's required is a historical sample of card transactions, an understanding of the business domain (such as vacation travel purchases), and the mastery of some novel analytic techniques related to discovery and scoring of transaction-based sequence analysis.

STRUCTURE OF EVENT SEQUENCES

Let's start by enumerating some fundamental concepts that describe the structure of event sequences. Sequences have **length**, equal to the number of events comprising that sequence; left-hand events are called **precedents**, while right-hand events are called **antecedents**. In a predictive analytics business domain, we are often interested in predicting the presence of a specific antecedent conditional on specific precedents; in the diagram below, shading is used to indicate two examples of sequences that don't occur in the sequence width, specifically A → A → A and A → C → C. It's up to the business analyst to decide which sequences are targeted, based on their business experience and objective.

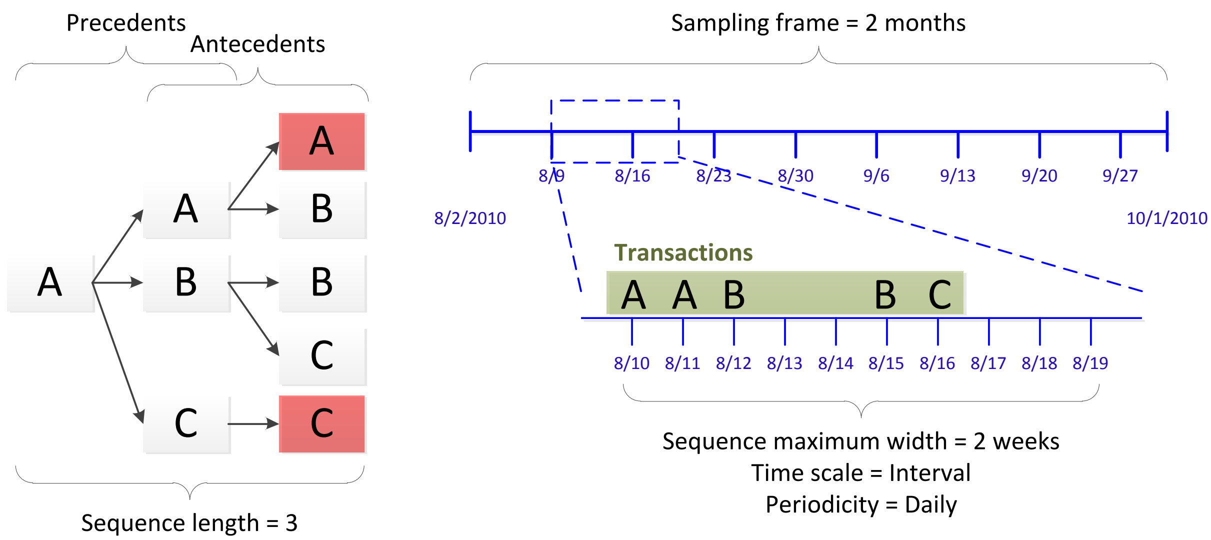


Figure 7: Structural components of event sequences

The time domain of an event sequence also contains some important concepts. The historical extent of the transaction history used in sequence discovery and scoring is called the **sampling frame**; in the diagram below the sampling frame runs between month-start August and month-start October of 2010. Any individual event sequence can be constrained to start and end within a defined timeframe called the **maximum width**; in the example below, the maximum width is two weeks long. The maximum width of an event sequence is a moving window within the sampling frame. **Time measurement** can be represented as either interval (which is traditional if the analysis relies on SAS date, time or datetime constants) or ordinal (if time is measured in more general terms). P**eriodicity** is the width of a single time unit, in this case daily. Each time unit can contain zero, one or more than one transaction, but in the example below, for ease of display, I've selected to only portray each day as containing zero or one transactions.

SEQUENCE ANALYSIS PROCESS

The process of sequence analysis is composed of eight steps: building the transaction table; filtering and aggregating transactions into events; cluster sampling; performing rule discovery via sequence model training; selecting meaningful sequence rules; scoring accounts on the selected sequence rules; deploying customer strategies tied to sequence rule scores; monitoring responses to the customer strategies and repeating this cycle.

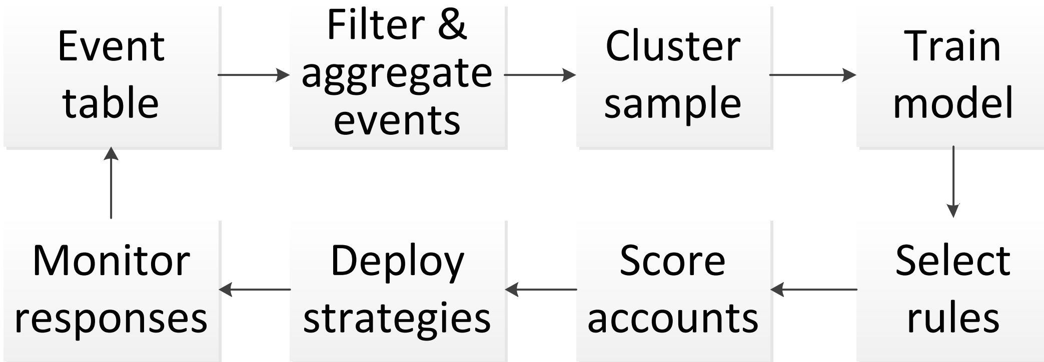


Figure 8: Stages of sequence analytics

1. **Building the event table**: This task is fairly straightforward. A transaction table consists of the customer entity ID, the transaction type, the timestamp, and any supporting information about the customer or transaction that might be used to stratify customers or refine transactions into events. In practice this table could contain tens or hundreds of millions of records, but it's relatively skinny, making it easy to handle. Furthermore if you are working with this table natively in a database, many of the tasks in step two below can be executed in-database via SAS pass-thru SQL.
2. **Filtering and aggregating transactions into events**: This task requires more planning, care, iterative refinement and data cleansing than most of the other steps in this process. Analogous to the 80/20 rule applied to most types of analytic endeavors, the filtering and aggregating of transaction records is the 80% of the arduous work that allows for the rule discovery and scoring to produce results as valuable as possible. While the list of transaction record data cleansing techniques I've used in practice is extensive and always growing, I'm happy to note that I'm building a course to be offered by SAS Institute where we will spend the better part of the day reviewing these techniques. This list of techniques includes identifying and removing duplicates, consolidating similar types of transactions, removing over-dominant and rare events, enriching the transaction file for left-hand and right-hand events of the greatest interest, developing transaction hierarchies, stratifying for predetermined clusters of customers, insertion of non-events, and conversion of events into customer behavior states.
3. **Cluster sampling**: This task is pretty simple; as with other predictive models, the analyst need not work with the entire universe of records in the transaction history in order to develop a decent sequence analysis. However, a random sample approach applied at the transaction level is not the desirable approach when working with transactions within customers. A variant of stratified sampling, called cluster sampling, is the appropriate method. Cluster sampling is used when the analyst wants to select a sample of higher-level entities, and then select all lower-level items that each sampled entity owns or is associated with. In population studies, cluster sampling might be used to select a sample of zip codes, and then select all households within sampled zip codes. In sequence analytics, the analyst should select a sample of customers, and then select all transactions associated with sampled customers. This task can be executed via hand written SQL, the SAS SURVEYSELECT procedure (starting with SAS 9.2 and later releases), and in SAS Enterprise Miner software, among other approaches.
4. **Performing rule discovery via sequence model training**: In practice, I have always relied on SAS Enterprise Miner software to conduct this task. The SAS Enterprise Miner Associations / Sequences node was purpose-built to rapidly discover and select data-driven event sequences from large transaction tables meeting user-defined constraints. These user-defined constraints include the length of the sequence rule chain, the maximum sequence width, the time measurement width, the minimum meaningful incidence rate of the sequence, and the maximum number of sequence rules to export for scoring. These features are essential as they allow the user a great deal of power in tuning the sequence discovery engine to identify the first layer of rule attributes. However, I've often found a second level of rule filtering driven by the analyst’s careful scanning and further filtering of rules is recommended. This is described in the next step.

5. **Selecting rules**: The Enterprise Miner tool also supports interactive filtering on a generated set of rules via an interactive dialog window; this is particularly useful when the analyst knows in advance whether a meaningful rule will always contains a certain left-hand side or right-hand side event type. For instance, if the purpose of sequence discovery is to augment and account delinquency predictive classification model, then it's perfectly acceptable and recommended for the onset of delinquency to be represented as an event in the transaction table; at that point the analyst should filter the generated set of rules for those that contain onset of delinquency in the right-hand side of the sequence chain. This approach enriches the generated list of sequence rules for those of direct interest for the subsequent predictive model.
6. **Scoring accounts**: As with many other types of predictive analysis, a model should be able to satisfy two important criteria: it should faithfully represent the process that generated the historical data, and it should be usable for predicting the outcome of the process that generates yet-to-be-recorded data. This latter step is often referred to as scoring. Many analysts familiar with Market Basket or sequence analysis might not be aware that such analyses can also be used to score future sets of customers for the presence of a set of association or sequence rules. SAS Enterprise Miner produces scoring code for association and sequence rules, and in recent releases of the software recommends the use of the MBSCORE procedure. It has been my observation that this approach, while powerful and simple to use, lacks a few key capabilities: it doesn't identify the date of onset of a sequence rule for an individual customer; it doesn't allow the user to input their own sequence rules for scoring (such as in the case of a disassociation rule), and as a SAS procedure, the analyst might be constrained from running the MBSCORE procedure on a production environment. To meet these challenges, I wrote a SAS data step routine that performs sequence scoring that meets all these requirements and performs remarkably fast from a computational perspective. Attendees of the SAS-sponsored course I will be teaching on sequence analytics will receive a copy of this scoring routine; other interested parties are welcome to contact me to request this approach for their own use.
7. **Deploying customer strategies tied to sequence rule scores**: Sequence scores at the customer entity level can provide a source of significant predictive lift for classification models used in a variety of customer strategies, including marketing, credit risk, collections and compliance analytics. Later in this paper, I describe some case studies where traditional predictive classification models have been augmented with sequence scores with remarkable incremental results. Furthermore, sequence scores can be used as attributes for triggered interactions, for eligibility rules in assigning customer treatments, and potentially in the dynamic treatment allocation to customers in an on-demand setting (e.g. by converting a sequence scoring routine into a web services-enabled SAS stored process).
8. **Monitor responses**: Most direct marketers, credit risk analysts and other decision scientists concerned with the performance of customer strategies, and the contribution of analytic models towards that performance, are familiar with transaction tables that monitor allocation in delivery of customer treatment and record the disposition of that treatment. In direct marketing, these tables are called contact history and response history. In collections, these tables are called activity and disposition. No matter the name, the analyst should be keenly interested in comparing contact and response rates for strategies augmented by sequence scores. Even more interesting, it is recommended that contact and response ought to be treated just like any other kind of transaction and included in sequence discovery, to enable the measurement of dynamic interaction between the firm and the customer as a type of event. This approach enables the refinement of contact strategies and the further enrichment of transaction event history tables for use in the next round of sequence rule development.

SEQUENCE RULE DEVELOPMENT, SCORING AND MODEL ENHANCEMENT USING SAS ENTERPRISE MINER

This section walks through the process of sequence rule development scoring model enhancement, featuring SAS Enterprise Miner software for nearly all the steps involved. The diagram below summarizes some of the critical components related to sequence discovery. In this diagram, the screen capture of the transaction history table (item 1) depicts loan payment history for an account numbered 36044 across various months on book (MOB 21 to 29); this account bounces back and forth across a series of delinquency bins from current to 60 days past due, and once to more than 60 days past due. In this analysis will use the account number to represent customer entity ID, the months on book (MOB) as an ordinal measure of time, and delinquency stage as the transaction event.

Based on auto loan payment & delinquency history

1 Loan payment history

ROBIN.MOB

	rowno...	MOB	DLQSTAGE	amount	delq_deldays
22	36044	21	0	$535.77	0
23	36044	21	1	$0.00	29
24	36044	22	2	$0.00	32
25	36044	22	0	$535.77	0
26	36044	22	1	$0.00	29
27	36044	23	2	$0.00	31
28	36044	23	0	$535.77	0
29	36044	23	0	$90.00	0
30	36044	24	1	$0.00	29
31	36044	25	2	$0.00	32
32	36044	25	0	$1,601.31	0
33	36044	27	1	$0.00	29
34	36044	28	2	$0.00	31
35	36044	29	3	$0.00	62
36	36044	29	0	$374.00	0

2 Enterprise Miner model diagram

3 Sequence rules tied to delinquency stage

Rules Table

Chain Length	Transaction Count	Support(%)	Confidence(%)	Rule	Chain Item 1	Chain Item 2
2	5636	96.21	96.24	0 ==> 0	0	0
2	1984	33.87	33.88	0 ==> 1	0	1
2	1542	26.32	68.05	1 ==> 0	1	0
2	1486	25.37	65.58	1 ==> 2	1	2
2	1369	23.37	60.41	1 ==> 1	1	1
2	1020	17.41	68.55	2 ==> 3	2	3

4 Sequence rules scored to each account

EMWS7.Assoc_TRAIN

	rown...	0 ==> 0	0 ==> 1	1 ==> 0	1 ==> 2	0 ==> 2	0 ==> 0 ==> 0	2 ==> 0
1	1439...	1	1	1	1	1	0	1
2	9440...	1	0	1	1	0	1	1
3	1052...	1	0	0	1	0	0	0
4	1257...	1	0	1	1	0	1	1
5	1583...	0	1	1	1	1	0	1
6	1980...	1	0	1	0	0	1	0
7	2472...	1	0	1	0	0	1	1

Figure 9: Sequence discovery components developed using SAS Enterprise Miner

An Enterprise Miner diagram (item 2) consists of three nodes, starting with the data source node (labeled “MOB”), a cluster sampling node and the sequences node. When the sequences task produces its results, it produces a rules table (item 3) that consists of one row per sequence rule, and attributes characterizing the rule, such as the rules’ confidence, support, and chain items.

Enterprise Miner can export a pivoted account by rule table for the sample of accounts used in sequence discovery, not accounting for multiple slices of the transaction window in the historical sampling frame (item 4): in this view a rule has occurred once or more than once (indicated by “1”) or not at all (indicated by “0”).

Enterprise Miner also provides a useful post-discovery rule filtering tool, depicted in the screen capture below. Starting at the bottom left corner of the window, the analyst has built a rule filtering search for sequence chains whose right-hand chain element contains the transaction type string ”open_loan”. The screen immediately above the search phrase indicates a list of the sequence rules meeting that criterion. The constellation diagram at the upper right corner of the window visually depicts the other chain items with which the “open_loan” event is associated, with measures of confidence and support indicated by arc and node attributes such as color and size. Most sequence analyses tend to stop at this point, treating sequences as interesting, descriptive observations of sequences that have occurred historically. However, my strong recommendation is to move forward into sequence scoring and classification model enhancement.

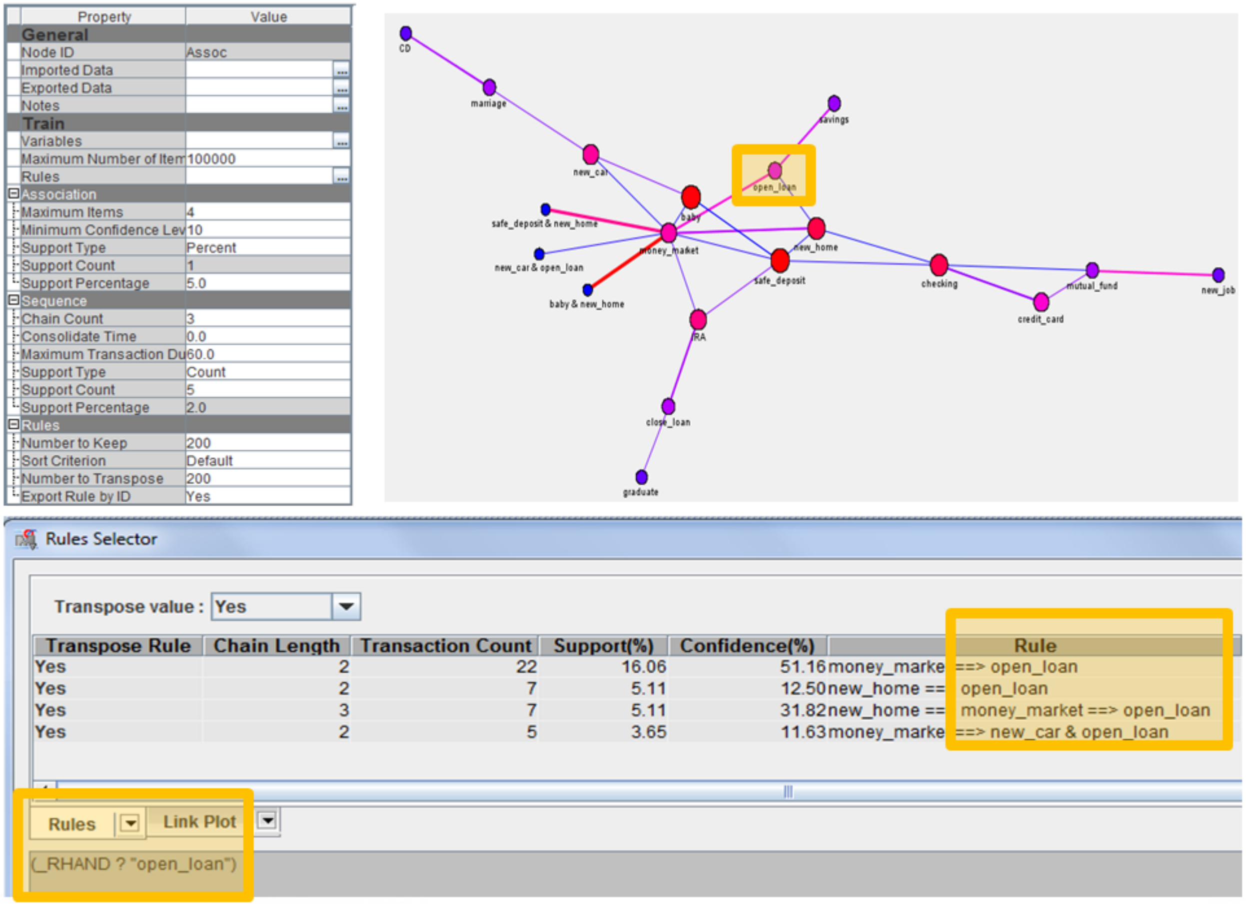
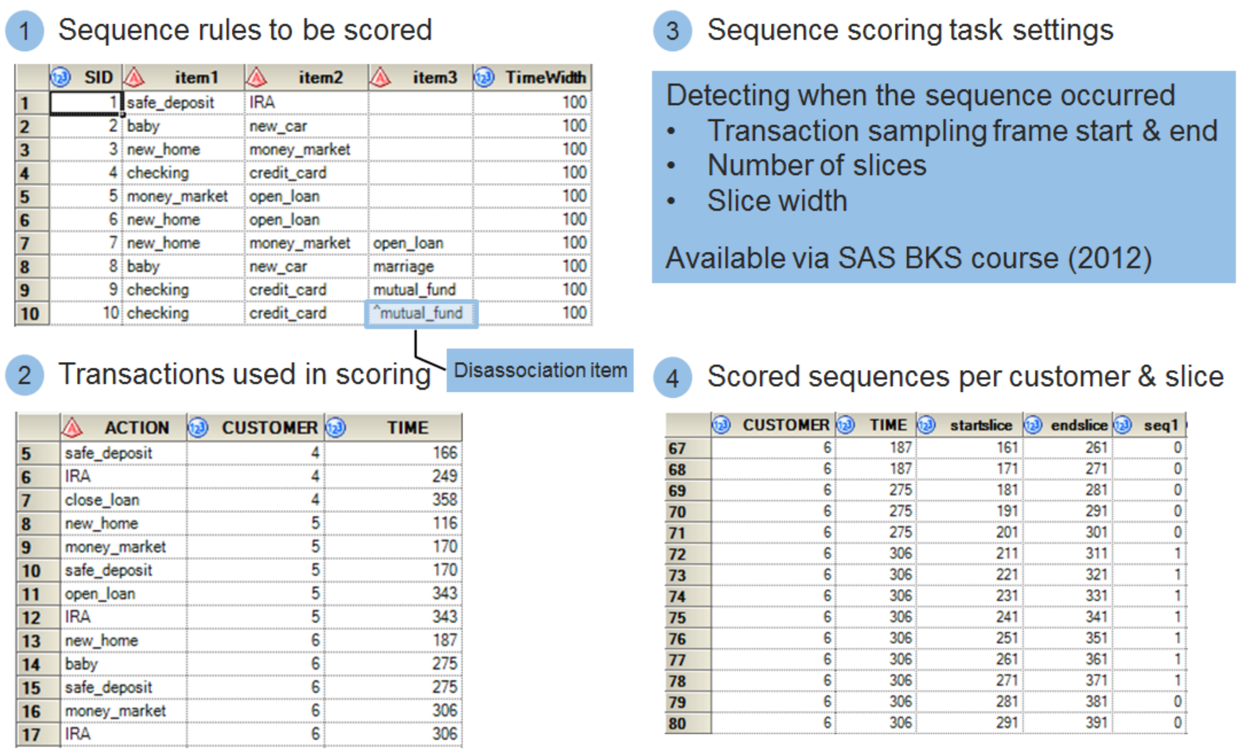


Figure 10: Rule filtering example

The next step is to score each customer in the original, non-sampled transaction history file, or perhaps a brand-new transaction history file selected for other customers, a different historical sampling frame, or both. In the list of sequence rules to be scored file (item 1), each record represents a sequence rule selected by the analyst, either from the Enterprise Miner-generated rules table, or rules that the analyst has decided are worthy of scoring whether they were discovered by Enterprise Miner or not. In each record in this list, a simple index for the sequence ID (“SID”) is accompanied by the items in the sequence chain, and the width of the moving time slice that will scan through the transaction history to be scored. Note that disassociations, indicating a sequence chain that includes a chain element representing the absence of a transaction type, is also supported by the sequence scoring mechanism I have developed. This is noted in the record for SID 10.



	SID	item1	item2	item3	TimeWidth
1	1	safe_deposit	IRA		100
2	2	baby	new_car		100
3	3	new_home	money_market		100
4	4	checking	credit_card		100
5	5	money_market	open_loan		100
6	6	new_home	open_loan		100
7	7	new_home	money_market	open_loan	100
8	8	baby	new_car	marriage	100
9	9	checking	credit_card	mutual_fund	100
10	10	checking	credit_card	^mutual_fund	100

	ACTION	CUSTOMER	TIME
5	safe_deposit	4	166
6	IRA	4	249
7	close_loan	4	358
8	new_home	5	116
9	money_market	5	170
10	safe_deposit	5	170
11	open_loan	5	343
12	IRA	5	343
13	new_home	6	187
14	baby	6	275
15	safe_deposit	6	275
16	money_market	6	306
17	IRA	6	306

	CUSTOMER	TIME	startslice	endslice	seq1
67	6	187	161	261	0
68	6	187	171	271	0
69	6	275	181	281	0
70	6	275	191	291	0
71	6	275	201	301	0
72	6	306	211	311	1
73	6	306	221	321	1
74	6	306	231	331	1
75	6	306	241	341	1
76	6	306	251	351	1
77	6	306	261	361	1
78	6	306	271	371	1
79	6	306	281	381	0
80	6	306	291	391	0

Figure 11: Sequence rule scoring components

The transaction history table to be scored contains one row for each customer and each transaction type, as well as the time in which the transaction occurred (item 2). The scoring routine is configurable by the analyst to indicate the start and end point of the transaction sampling frame, the number of slices to score and the width of each slice (item 3). Finally, the scored sequence table contains a unique row for each account at a specified point in the sampling frame, and each column represents the presence of a historical event occurring for that account at that point in the sampling frame. A “1” indicates the presence of the sequence event, and “0” indicates its absence.

This final screen capture depicts an Enterprise Miner process flow diagram representing a baseline model that does not use sequence scores, and compares it with a challenger model that does use sequence scores. The sequence scores file is filtered to select a single time slice, is merged with the model development table by customer ID, and their the baseline and challenger models are estimated using to alternative paths in the process flow diagram. The baseline model is estimated using the logistic regression node at the top of the diagram, while the challenger model is estimated in the bottom path. Variable clustering is used as an information reduction method to select exemplar sequence score variables so that highly collinear sequences won't swamp the resulting challenger regression model. The charts at the bottom of the screen capture depict three pairs of diagnostic charts: the comparative ROC chart, gains chart, and K-S chart (i.e., each pair of charts shows model performance for the training and validation partitions, respectively). The challenger model with its sequence score predictor variables outperforms the baseline model on all counts.

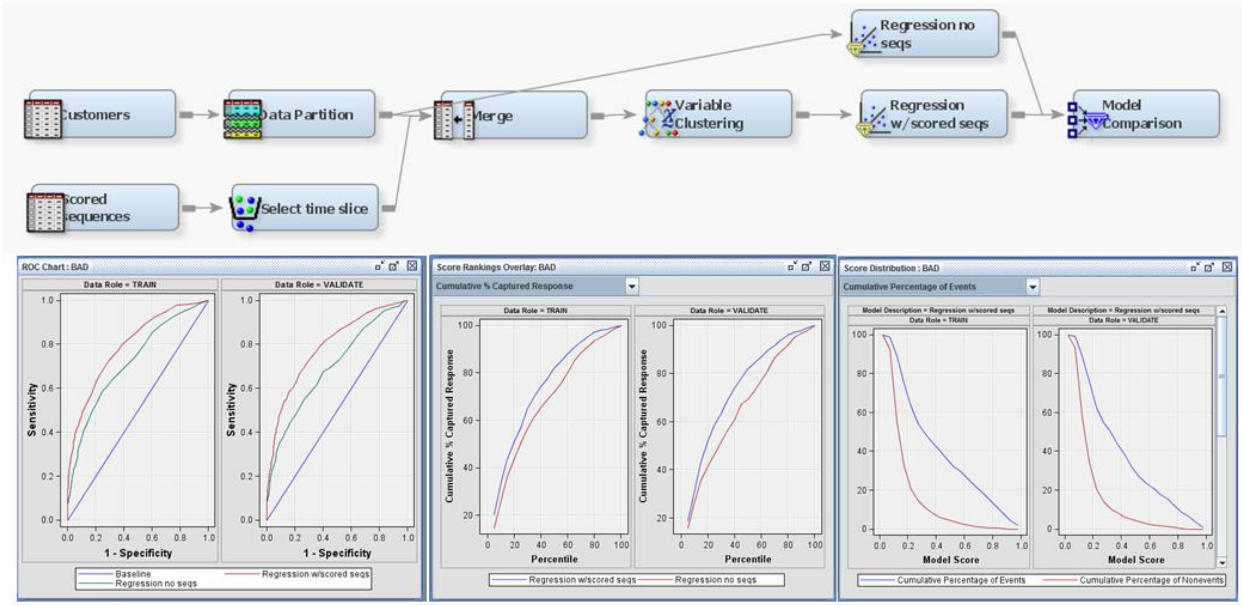


Figure 12: Comparison of classification model results using/not using sequence scores as predictors

CASE STUDIES

This paper ends with four brief real-world case studies conducted by the author with some of his North American retail financial services clients, in the payments, lending, brokerage and credit card sectors respectively.

We ran the campaign targeting, scoring and performance measurement process for a leading payment processor's targeted merchant mail campaign over an 18-month period. We developed transaction-based behavioral segmentation and scoring routines to maximize incremental credit card spend that routinely outperformed standard targeting approaches such as category-specific spending indices. In one such campaign, our targeted segments delivered a $6.9 million projected incremental spend on a marketing budget of $1 million. A hallmark of the approach we developed for these campaigns was the characterization of purchase event sequence rules, some focused inside the merchant partner's category and others sampled from a wide variety of merchant categories. In side-by-side tests with non-sequence model campaign cells, the sequence model campaign cells routinely outperformed.

We developed a mathematical optimization routine for the auto loan servicing department of a retail bank to maximize the risk-weighted returns on collections activities by assigning the ideal delinquency & recovery strategy to each of roughly 170,000 accounts in the auto loan portfolio. Key deliverables included behavioral segmentation and the construction of over 20 delinquency & collections scorecards and a collateral valuation model. The optimized strategy assignment routine accounted for variable costs and resource capacity per collection strategy, the eligibility of strategies per account given the maturity of the loan and the collateral book-to-value. A key contribution to the regression model scores used to assign collection treatments to debtors were sequence scores based on payment history and collection activity history, in particular because those model scores helped target the proper timing for assignment of treatments based on what had occurred between the lender and the debtor in recent periods.

We developed a behavioral segmentation strategy and a pair of predictive scorecards for a money management firm that sought to better target their non-captive financial advisors for sales outreach leading to increased assets under management. The segmentation was comprised of advisor attributes, past sales and redemption patterns and transaction event sequences; the predictive scorecards focused on increasing the targeted response rate by scoring on transaction volumes and segmentation-based factors. The increased response rate from the scorecard exceeded business-as-usual predictive accuracy by over 20%.

In the credit card issuing sector, we augmented an existing predictive scorecard for a credit card issuer to improve the leading-indicator detection rate of active-spending cardholders to migrate into inactive behavior. We developed novel predictors tied to spend transaction behavior and transaction sequence models that increased the migration event prediction rate by over 100% on transactor accounts and by more than 15% on revolver accounts in the top decile of targeted accounts. A novel part of our sequence scoring approach was to summarize cardholder behavior to a behavior state characterizing higher than expected, lower-than-expected, or zero level activity in each semimonthly (i.e., 15-day) period, and to construct sequences on these behavior states. Our predictors also revealed the likely behavioral drivers that could be used to more effectively stratify and position the contact strategy with accounts selected for intervention.

CONCLUSION

In this paper, we've introduced the concept of event sequences defined the structure of sequence analytics, walked through a practical example using some of the tools of the trade, and proved our point with four case studies drawn from real-world examples in the financial services industry. Anyone interested in learning more about event sequence analytics is encouraged to register for the soon-to-be-released SAS-sponsored course on financial transaction sequence modeling, written and taught by the author. It's my hope that this paper stimulates interest and excitement in the predictive analytics community for this fascinating and powerful analytic approach.

CONTACT INFORMATION

Name:	Robin E. Way, Jr.
Enterprise:	Corios LLC
Address:	133 SW Second Avenue, Suite 202
City, State ZIP:	Portland OR 97204
Work Phone:	503.295.1685
E-mail:	rway@coriosgroup.com
Web:	http://coriosgroup.com