Predictive Analytics in Big Data

Bel G Raggad, Seidenberg School of CS & IS, Pace U, New York

Abstract - Data has achieved a size beyond which predictive analysis has become costlier and more complex. Despite this dead end, we still continue spending billions of dollars amassing cosmic volumes of data without planning any feasible approaches to gain the long awaited business value for which we trusted data to lead to. That said, we know that unless we manage to dig deep in these data to understand our decision parameters, we will still lack the predictive power needed to make a sound decision. In order to plan our decision, we first needed to construct probability distributions for all relevant decision parameters and prepare to apply available decision theoretic methods. Without the sought probabilistic decision support, there will be no sound method to act.

We propose a predictive analytic model, using Smet's Transferred Belief Model to construct probability distributions for all our decision parameters based on sufficient representative subsets of data throughout the big data depots. Once this probabilistic framework is in place, available decision theory models can be applied.

A fictitious numeric example is processed to demonstrate the working of our predictive analytic model.

Keywords - Predictive analytics, Transferred Belief Model, Dempster and Shafer theory, big data, probability distributions.

Introduction

The literature started discussing big data at a fast pace. Unfortunately most of the reported studies dealt with proposing data analytic models that are more appropriate for traditional statistics and data mining. [1, 4, 5] Some other studies focused on storage management which is an important topic in the big data field. So we are still in need for tools to manage business value based on data that is at risk to grow so large that they may become impossible to work with using available traditional data analytic approaches. [4, 5]

Big data sizes have reached the petabytes in one data depot and the entire big data stores have become just too creepy to even start exploring them. We just cannot any longer know how we can feasibly continue the collection, the storage, the search, and the planning of any gain of business value. [4, 5] We certainly see great analytics tools proposed to process large data samples and they may reveal great business benefits, but these techniques are still very local and do not take advantage of the properties of big data, especially when its volumes expand, its velocity increase, and it variability become overwhelming. [1, 4]

We are still attempting to seek sound methodologies capable to analyze larger and more complex data sets with the ability to efficiently manage real time and live data streams. We are still in need for big data management platforms, new data architectures that ease the development and the adoption of powerful analytical methods and tools. Any big data analytic approaches should take into account of the main big data properties, including its first three V's: volume, variety, and velocity. The volume property addresses the massive scale and growth of unstructured data beyond traditional storage and analytical solutions. The variety property addresses the traditional data management processes beyond the heterogeneity of the data. [1, 11] The velocity property addresses the speed of production of data, especially for real-time data, with the objective of immediate business advantage. [3] The literature also advanced other studies on big data that proposed other V's [1, 11] related to data inconsistency, incompleteness, ambiguity, latency, deception, and many other data quality properties. [1, 11]

In this paper intends to extend the literature to propose an analytical approach to construct probability distributions for decision parameters relevant to a decision problem using available big data resources. Sound decision theoretic techniques cannot be carried out if the needed probability distributions on decision parameters cannot be obtained. [2] Decision theory is needed to study decision problems and to fit them into a framework where the decision maker can have all the techniques to manage all types of uncertainties before a decision is made. [2] The relevant decision parameters have to be studied based on their probability distributions and the expected outcomes of available alternative decision actions. The decision maker needs to understand the actions consequences when taken by using utility theory in valuing all types of outcomes. Bayesian decision theory offers a statistical platform that provides the techniques to quantify the tradeoffs between various outcomes, based on probabilities and on expected values of costs. That said, nothing of this can be achieved if those probability distributions on the decision parameters cannot be produced despite all the data accumulated in costly ways.

Probabilistic model based on TBM

We intend to develop an analytical model to examine sufficient data in a big data depot on decision parameters that are relevant to a given decision problem for the purpose of producing the decision support information needed to take action. For this purpose, we consider a big data depot with known structure that is feasibly made available to our data analytics team. That is, let us assume that we now have on hand an extracted smaller big data depot ϴ with a known structure, say, $\{\{a_{ij}\}i=1,M; j=1,N\}$, M,N: sufficiently large} where a_{ij} is a (simple) data values in $A_1 \times \ldots \times A_N$. Also let $<q_{i1+1}, \ldots, q_{iK+M1}>$ a sequence of K integers between 1 and $M-M_1$ such that $|q_k-q_{k-1}|>M_1$ for $1<k<K$ that serve as pointers needed to navigate the big data to selected our data subsets D_j, j=1,N that we process using the TBM model to produce probability distributions.

Construction of belief structures on selected data subsets

Given a data subset D_{qk}, for every k, $1 \leq k \leq K$, we can construct a belief structure by considering the frame of discernment $\Omega=\prod_{i=1,N} 2^{Ai}$ as the Cartesian product of all power sets of data attributes on the selected data subset. Because the big data is structured, this frame of discernment is the same foe all data sets. We propose the following steps as in Wang [12] that can be followed for the construction of belief structures of the selected data subsets:

Step 1: Define the generic hyper power set of selected hyper attributes from the selected big data sub-depot

Step 2: For every subset g in Ω, compute $s_{Dqk}(g)= \sum_{e \text{ in } g} |\{d \text{ in }_{Dqk} \text{ such that } e \text{ in } d\}|$, the support of g in the data subset Dqk , for every k, $1 \leq k \leq K$.

Step 3: Compute the belief structure:

m_{Dqk}: $2^{\Pi_{i=1,N1}Ai}m \rightarrow [0\ 1]$
$m_{Dqk}(e) = s_{Dqk}(\{e\})/\ s_{Dqk}(G)$
$m_{Dqk}(\Omega) = 1$

Index i=1,M	**Selected big data subsets for TBM processing**					
$i=q_{i1}$						
		A_1		A_j		A_{N1}
	t_{qi1+1}	$a_{qi1+1,1}$		$a_{qi1+1,j}$		$a_{qi1+1,N1}$
	- - - -	- - - -	- - - -	- - - -	- - - -	- - - -
	t_{qi1+k}	$a_{qi1+k,1}$		$a_{qi1+k,j}$		$a_{qi1+k,N1}$
	- - - -	- - - -	- - - -	- - - -	- - - -	- - - -
	t_{qi1+M1}	$a_{qi1+M1,1}$		$a_{qi1+M1,j}$		$a_{qi1+M1,N1}$
- - - -	- - - - - - - - - - - -					
$i=q_{ik}$						
		A_1		A_j		A_{N1}
	t_{qik+1}	$a_{qik+1,1}$		$a_{qik+1,j}$		$a_{qik+1,N1}$
	- - - -	- - - -	- - - -	- - - -	- - - -	- - - -
	t_{qik+k}	$a_{qik+k,1}$		$a_{qik+k,j}$		$a_{qik+k,N1}$
	- - - -	- - - -	- - - -	- - - -	- - - -	- - - -
	t_{qik+M1}	$a_{qik+M1,1}$		$a_{qik+M1,j}$		$a_{qik+M1,N1}$
- - - -	- - - - - - - - - - - -					
$i=q_{iK}$						
		A_1		A_j		A_{N1}
	t_{qiK+1}	$a\ _{qiK+1,1}$		$a_{qiK+1,j}$		$a_{qiK+1,N1}$
	- - - -	- - - -	- - - -	- - - -	- - - -	- - - -
	t_{qiK+k}	$a\ _{qiK+k,1}$		$a_{qiK+k,j}$		$a_{qiK+k,N1}$
	- - - -	- - - -	- - - -	- - - -	- - - -	- - - -
	t_{qiK+M1}	$a\ _{qiK+M1,1}$		$a_{qiK+M1,j}$		$a_{qiK+M1,N1}$
- - - -	- - - - - - - - - - - -					

Fusing belief structures

We have K sources of evidence which are the hyper data subsets $\{D_{qk}\}_{k=1,K}$ that tell us about the behaviors of all focal hypertuples in the feasible space Ω. Before we can compute the basic belief assignment values associated with hypertuples in Ω we need to fuse all available pieces of evidence obtained from these sources. At this point, Depmster's Rule comes handy to apply to the K belief structures $\{m_{Dqk}\}_{k=1,K}$ for the production of the combined evidence m_D where D is the set of hyper data subsets $\{D_{qk}\}_{k=1,K}$.

Dempster's rule of combination is a rule for combining belief functions when these belief functions are based on independent sources of evidence. [9, 10]

Specifically, the combination is calculated from the two sets of masses *m*₁ and *m*₂ as follows:

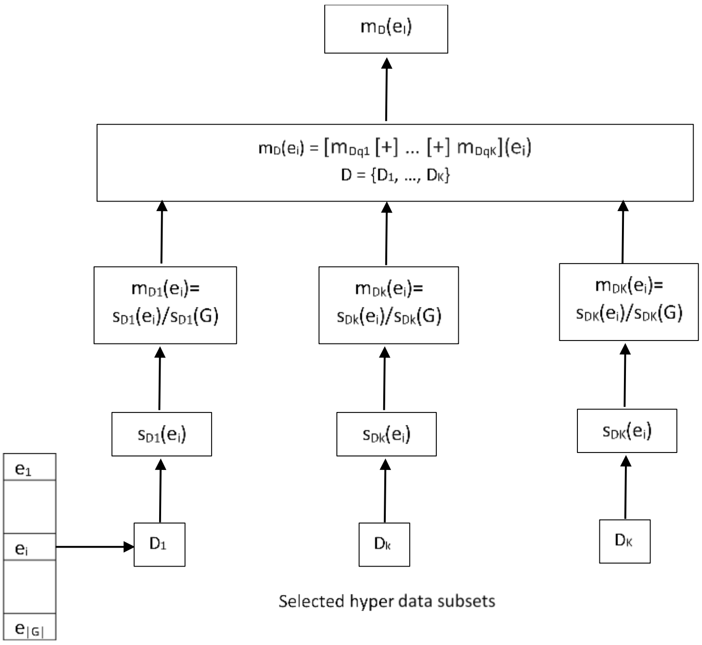
$m_{12}(\emptyset)=0$
$m_{12}(A)=\ [m_1(+)m_2](A) = 1/(1-k)\ \sum_{B\cap C=A\neq\emptyset} m_1(B)m_2(C)$, where :
$k= \sum_{B\cap C=\emptyset} m_1(B)m_2(C)$ is a measure of the amount of conflict between the two bba's.

After applying Dempster’s rule, we obtain the fused belief structure m_D as follows [12]:

$m_D(\varnothing)=0$
$m_D(e)= [m_{Dq1} [+] \dots [+] m_{DqK}](e) = 1/(1-k) \sum_{\cap k=1,K (xk)=e} \prod_{k=1,K} m_{Dqk}(x_k)$, where :
$k = \sum_{\cap k=1,K (xk)=\varnothing} \prod_{k=1,K} m_{Dqk}(x_k)$ is a measure of the amount of conflict between the K bba’s.

In order to avoid the computations to produce the fused belief structure using Dempster’s Rule can be very complex sometimes, including in this case, we are using some easier equations proposed by in Wang et al. (2007) as follows [12]:

$$m_D(e)= [m_{Dq1} [+] \dots [+] m_{DqK}](e) = \sum_{k=1,K} m_{Dqk}(e).s_{Dqk}(\Omega) / \sum_{k=1,K} s_{Dqk}(\Omega)$$


$m_D(e_i)$
$m_D(e_i) = [m_{Dq1} [+] \dots [+] m_{DqK}](e_i)$
$D = \{D_1, \dots, D_K\}$
$m_{D1}(e_i)= s_{D1}(e_i)/s_{D1}(G)$
$m_{Dk}(e_i)= s_{Dk}(e_i)/s_{Dk}(G)$
$m_{DK}(e_i)= s_{DK}(e_i)/s_{DK}(G)$
$s_{D1}(e_i)$
$s_{Dk}(e_i)$
$s_{DK}(e_i)$
e_1
e_i
$e_{|G|}$
D_1
D_k
D_K
Selected hyper data subsets

At this point, we have succeeded in constructing a belief structure for each of the K subsets we arbitrarily selected from the structured big data depot. We also showed how to fuse the belief structures using Dempster rule, and showed how to compute the basic belief assignment value (mass value) for every focal element in the feasible space G based on the hyper data subsets. We now then have accumulated all available evidence about the behaviors of all focal elements in the hyper data subset D. We now started to have an idea about the behavior of our decision parameters but not to the point where we can make a mathematically sound decision. In order to do so, we need to manage all the uncertainty associated with our decision domain G; and for this, we need to come up with a way to construct probability distributions, if possible, for all relevant decisions parameters in the feasible space. For this purpose, we propose applying Smets’ Transfer Belief Model (TBM). [8] This model is capable to produce approximative probability distributions for all our decision parameters, called the pignistic probabilities. Once we have them, we can then apply decision theory to act in a mathematically sound manner.

Smets’ Transfer Belief Model

Let us have a brief introduction of the Smets' Transferred Belief Model (TBM) [8]. The TBM consists of two steps: the credal model step and the pinistic model step. The reader may alternatively opt for Shafer's plausibility functions as a substitute to Smets' pignistic probabilities, as both techniques stem from the same belief structure and both add greater interpretability to the TBM. [6, 7, 8].

The design of the creedal step may be set to fully asserted evidence based on the selected subsets without accounting for any managerial judgment that may be sometimes relevant in some decision problems and without accounting for any certainty factors or discount factors associated with the evidence on hand. This means that the basic belief assignments expressing the uncertainty associated with the hyper data subsets' evidence remain fully asserted. The overall evidence on hand on our decision parameters has been accepted and expressed as a single belief structure as given above:

$m_D(\emptyset)=0$
$m_D(e)= [m_{Dq1} [+] \dots [+] m_{DqK}](e) = 1/(1-k) \sum_{\cap k=1,K\,(xk)=e} \prod_{k=1,K} m_{Dqk}(x_k)$, where :
$k = \sum_{\cap k=1,K\,(xk)=\emptyset} \prod_{k=1,K} m_{Dqk}(x_k)$ is a measure of the amount of conflict between the K bba's.

As mentioned earlier, even though we here demonstrate the pignistic model, another way may alternatively choose to compute Shafer's plausibility functions as a substitute to the pignistic probabilities. Smets' pignistic probabilities may be induced from the above belief structure as follows:

For any e in Ω:
$P_{Bet}(e) = \sum_{e\leq x} m_D(x).|x\cap e|/|x|$.

Numerical Example

The initial size of the power set of $A_1xA_2xA_3$ is $2^{|A1|}.2^{|A2|}.2^{|A3|}$ which is 8x8x4 = 256 any hypertuples. Let us assume that we are only concerned with the attributes A_1 and A_2, and we are hence only processing $2^{|A1|}.2^{|A2|}$=8x8=64 hypertuples.

For demonstration purposes, we will select K=3 simple data subsets of a fixed size equal to M1=5. The pointers are arbitrarily set at q_1=8, q_2=34, and q_3=53.

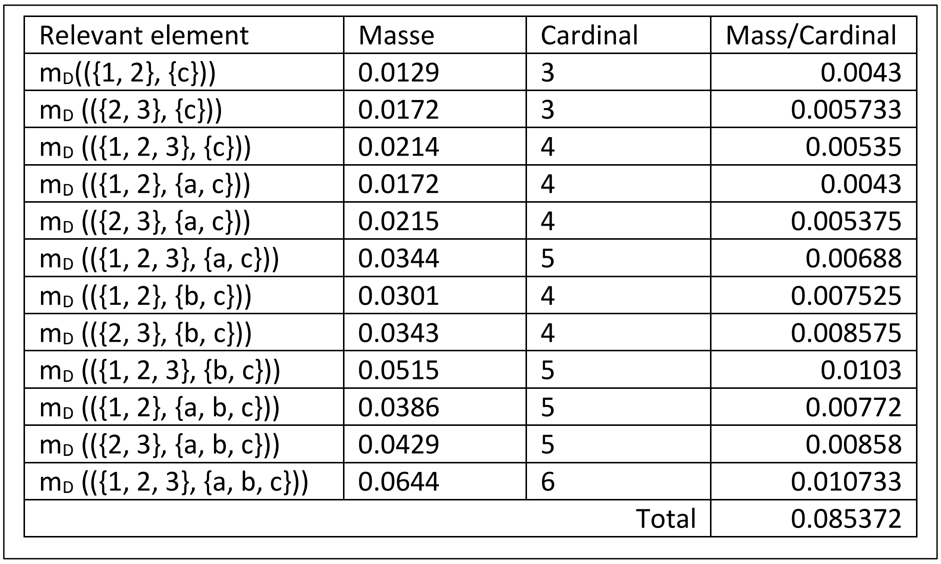
In order to apply common decision theory methods, we need to know the probability distributions of the two attributes A_1 and A_2 or at least the probability of their Cartesian product A_1xA_2.

In this example, the elements of A1xA2 for which we need probability distributions are only |A1|x|A2| = 3x3=9 elements which are {(1, a), (1,b), (1, c), (2, a), (2,b), (2, c), (3, a), (3,b), (3, c)}.

The pignistic probability for an element a = (a1, a2) in A1xA2 is given as follows:

$P_{Bet}(a) = \sum_{a\varepsilon x\leq A1xA2} m(x)/|x|$

For example, P_{Bet} ((2,c)) = m(({1, 2}, {c}))/|({1, 2}, {c})| + m(({2, 3}, {c}))/|({2, 3}, {c})| +
m(({1, 2, 3}, {c}))/|({1, 2, 3}, {c})| +
m(({1, 2}, {a, c}))/|({1, 2}, {a, c})| + m(({2, 3}, {a, c}))/|({2, 3}, {a, c})| +
m(({1, 2, 3}, {a, c}))/|({1, 2, 3}, {a, c})| +
m(({1, 2}, {b, c}))/|({1, 2}, {b, c})| + m(({2, 3}, {b, c}))/|({2, 3}, {b, c})| +
m(({1, 2, 3}, {b, c}))/|({1, 2, 3}, {b, c})| +
m(({1, 2}, {a, b, c}))/|({1, 2}, {a, b, c})| + m(({2, 3}, {a, b, c}))/|({2, 3}, {a, b, c})|
+ m(({1, 2, 3}, {a, b, c}))/|({1, 2, 3}, {a, b, c})|



Relevant element	Masse	Cardinal	Mass/Cardinal
m_D(({1, 2}, {c}))	0.0129	3	0.0043
m_D (({2, 3}, {c}))	0.0172	3	0.005733
m_D (({1, 2, 3}, {c}))	0.0214	4	0.00535
m_D (({1, 2}, {a, c}))	0.0172	4	0.0043
m_D (({2, 3}, {a, c}))	0.0215	4	0.005375
m_D (({1, 2, 3}, {a, c}))	0.0344	5	0.00688
m_D (({1, 2}, {b, c}))	0.0301	4	0.007525
m_D (({2, 3}, {b, c}))	0.0343	4	0.008575
m_D (({1, 2, 3}, {b, c}))	0.0515	5	0.0103
m_D (({1, 2}, {a, b, c}))	0.0386	5	0.00772
m_D (({2, 3}, {a, b, c}))	0.0429	5	0.00858
m_D (({1, 2, 3}, {a, b, c}))	0.0644	6	0.010733
		Total	0.085372

At this point we succeeded to construct the probability distribution of our decision parameters consisting of the Cartesian product of A_1 and A_2. Once this probability distribution is obtained we can then also compute the variance and standard deviation on the decision parameters and we can compute the expected values of the parameters and any other quantities on interest to the decision making process. Without the probability distributions of the decision parameters there will be no sound mechanism to apply available decision theoretic approaches.

Managerial implications

Decision theory is the study of decision problems and preparing them in a framework where one has all the techniques to manage all types of uncertainties before a decision is made. All relevant decision parameters have to be studied based on their probability distributions and decision outcomes. An example of the methods one needs to understand the consequences is the use of utility theory in valuing all types of outcomes. Bayesian decision theory is a statistical platform that provides the techniques to quantify the tradeoff between various outcomes, based of probabilities and expected values of costs. That said, nothing of the above can be accomplished if those probability distributions on the decision parameters cannot be produced despite all the abundances of data we accumulated in costly ways.

It is then about time that managers redeem the benefits they long awaited from accumulated data. The traditional statistical methodologies, including data mining, will not work, given the size of data, its variety, and its velocity. The predictive analytic model we proposed is useful for those managers who have the data but cannot generate the business value needed to redefine marketing strategies, increase sales of services and goods, manage rivalry, and achieve a lasting business continuity.

We provided a simple numerical example to demonstrate how our proposed model works. The size of the data sets can be adjusted to your company needs and also the number of the data subsets and their dispersion in the big data depots to achieve maximum representativeness. The use of a random generator to set the pointers where data subsets can be started will also improve the representativeness of the data subsets. The predictive analytic process can also set to be iterative to achieve convergence. This convergence can be achieved by imposing a stopping criterion where the iterative process stops as soon as we notice stable probability distributions for the decision parameters, i.e., no significance change in the probability distributions have been observed in the last batch of iterations.

Frame $2^{A_1 x A_2}$: $S_{D1}(G)$=79, $S_{D2}(G)$=79, $S_{D3}(G)$=75 –Continued-

Index	A_1	A_2	$S_{D1}(e)$	$S_{D2}(e)$	$S_{D3}(e)$	$m_{D1}(e)$	$m_{D2}(e)$	$m_{D3}(e)$	$m_{D1,D2,D3}(e)$
25	{1, 2}	{a, b}	3	1	3	0.0380	0.0127	0.0400	0.0301
26	{1, 3}	{a, b}	2	3	2	0.0253	0.0380	0.0267	0.0301
27	{2, 3}	{a, b}	1	2	3	0.0127	0.0253	0.0400	0.0258
28	{1, 2, 3}	{a, b}	3	3	4	0.0380	0.0380	0.0533	0.0429
29	{1}	{a, c}	2	0	0	0.0253	0.0000	0.0000	0.0086
30	{2}	{a, c}	1	1	0	0.0127	0.0127	0.0000	0.0086
31	{3}	{a, c}	0	2	1	0.0000	0.0253	0.0133	0.0129
32	{1, 2}	{a, c}	3	1	0	0.0380	0.0127	0.0000	0.0172
33	{1, 3}	{a, c}	2	1	1	0.0253	0.0127	0.0133	0.0172
34	{2, 3}	{a, c}	1	3	1	0.0127	0.0380	0.0133	0.0215
35	{1, 2, 3}	{a, c}	3	3	2	0.0380	0.0380	0.0267	0.0344
36	{1}	[b, c}	2	1	1	0.0253	0.0127	0.0133	0.0172
37	{2}	[b, c}	2	1	1	0.0253	0.0127	0.0133	0.0172
38	{3}	[b, c}	0	2	2	0.0000	0.0253	0.0267	0.0172
39	{1, 2}	[b, c}	3	2	2	0.0380	0.0253	0.0267	0.0301
40	{1, 3}	[b, c}	2	3	3	0.0253	0.0380	0.0400	0.0343
41	{2, 3}	[b, c}	2	3	3	0.0253	0.0380	0.0400	0.0343
42	{1, 2, 3}	[b, c}	4	4	4	0.0506	0.0506	0.0533	0.0515
43	{1}	{a, b, c}	3	1	1	0.0380	0.0127	0.0133	0.0215
44	{2}	{a, b, c}	2	1	2	0.0253	0.0127	0.0267	0.0215
45	{3}	{a, b, c}	0	3	2	0.0000	0.0380	0.0267	0.0215
46	{1, 2}	{a, b, c}	5	2	2	0.0633	0.0253	0.0267	0.0386
47	{1, 3}	{a, b, c}	3	4	3	0.0380	0.0506	0.0400	0.0429
48	{2, 3}	{a, b, c}	2	4	4	0.0253	0.0506	0.0533	0.0429
49	{1, 2, 3}	{a, b, c}	5	5	5	0.0633	0.0633	0.0667	0.0644

Frame $2^{A_1 x A_2}$: $S_{D1}(G)$=79, $S_{D2}(G)$=79, $S_{D3}(G)$=75

Index	A_1	A_2	$S_{D1}(e)$	$S_{D2}(e)$	$S_{D3}(e)$	$m_{D1}(e)$	$m_{D2}(e)$	$m_{D3}(e)$	$m_{D1,D2,D3}(e)$
1	{1}	{a}	0	0	0	0.0127	0.0000	0.0000	0.0043
2	{2}	{a}	0	0	0	0.0000	0.0000	0.0000	0.0000
3	{3}	{a}	0	1	0	0.0000	0.0127	0.0000	0.0043
4	{1, 2}	{a}	1	0	1	0.0127	0.0000	0.0133	0.0086
5	{1, 3}	{a}	0	1	0	0.0127	0.0127	0.0000	0.0086
6	{2, 3}	{a}	1	1	1	0.0000	0.0127	0.0133	0.0086
7	{1, 2, 3}	{a}	1	1	1	0.0127	0.0127	0.0133	0.0129
8	{1}	{b}	1	1	1	0.0127	0.0127	0.0133	0.0129
9	{2}	{b}	1	0	1	0.0127	0.0000	0.0133	0.0086
10	{3}	{b}	1	1	1	0.0000	0.0127	0.0133	0.0086
11	{1, 2}	{b}	2	1	2	0.0253	0.0127	0.0267	0.0215
12	{1, 3}	{b}	2	2	2	0.0127	0.0253	0.0267	0.0215
13	{2, 3}	{b}	2	1	2	0.0127	0.0127	0.0267	0.0172
14	{1, 2, 3}	{b}	3	2	3	0.0253	0.0253	0.0400	0.0300
15	{1}	{c}	0	0	0	0.0127	0.0000	0.0000	0.0043
16	{2}	{c}	0	1	0	0.0127	0.0127	0.0000	0.0086
17	{3}	{c}	1	1	1	0.0000	0.0127	0.0133	0.0086
18	{1, 2}	{c}	0	1	0	0.0253	0.0127	0.0000	0.0129
19	{1, 3}	{c}	1	1	1	0.0127	0.0127	0.0133	0.0129
20	{2, 3}	{c}	1	2	1	0.0127	0.0253	0.0133	0.0172
21	{1, 2, 3}	{c}	1	2	1	0.0253	0.0253	0.0133	0.0214
22	{1}	{a, b}	1	1	1	0.0253	0.0127	0.0133	0.0172
23	{2}	{a, b}	2	0	2	0.0127	0.0000	0.0267	0.0129
24	{3}	{a, b}	1	2	1	0.0000	0.0253	0.0133	0.0129

Conclusion

We discussed the costly effort we make in amassing data. This effort has not been matched with an equal effort of analyzing the data and build the predictive power needed for our decision processes.

Bayesian decision theory provides great techniques to quantify the tradeoff between various outcomes, based of probabilities and expected values of costs. Unfortunately, none of these benefits can be accomplished if those probability distributions on the decision parameters cannot be produced despite all the abundances of data we accumulated in costly ways.

We proposed a predictive analytic model, using Smet's Transferred Belief Model to construct probability distributions for all our decision parameters based on sufficient representative subsets of data throughout the big data depots. Once this probabilistic framework is in place, available decision theory models can be applied.

We provided a simple numerical example to demonstrate how our proposed model works. The size of the data sets can be adjusted to satisfy company needs and also the number of the data subsets and their dispersion in the big data depots to achieve maximum representativeness. This model can be expanded, in a future research project, by employing a random generator to set the pointers where data subsets can be started. The predictive analytic process can also be set to be iterative to achieve convergence. This convergence can be achieved by imposing a stopping criterion where the iterative process stops as soon as we obtain stable probability distributions for the decision parameters (i.e., no significance change in the probability distributions have been observed in the last batch of iterations).

References

1. Elgendy, A. and A. Elragal, Big Data Analytics: A Literature Review Paper, P. Perner (Ed.): ICDM 2014, LNAI 8557, pp. 214–227, 2014.

2. Hens, T. and M.O. Rieger, Financial Economics, Springer Texts in Business and Economics, Springer-Verlag Berlin Heidelberg 2016.

3. Intel, a vision of big data, https://www.intel.com/content/dam/www/public/us/en/documents/reports/intel-corp-big-data-policy-position-paper.pdf, 2017].

4. Kubick, W.R.: Big Data, Information and Meaning. In: Clinical Trial Insights, pp. 26–28 (2012).

5. Russom, P.: Big Data Analytics. In: TDWI Best Practices Report, pp. 1–40 (2011)

6. Shafer, G., A Mathematical Theory of Evidence, Princeton University Press, 1976.

7. Shafer, G. and R. Srivastava, "The Bayesian and Belief-Function Formalisms: A General Perspective for Auditing," Auditing: A Journal of Practice & Theory, Vol. 9 Supplement, pp 110-137, 1990.

8. Smets, P. The Combination of evidence in the transferable belief model. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 12(5), pp 447-458, 1990.

9. Srivastava, R. P., and Mock, T. Why we should consider belief functions in auditing research and practice*. The Auditor's Report*, 28(2), pp 58-65, 2005.

10. Srivastava, R. P., and Liu, L. Applications of belief functions in business decisions: A review. *Information Systems Frontiers* 5(4), pp 359-378*, 2003.*

11. TechAmerica: Demystifying Big Data: A Practical Guide to Transforming the Business of Government. In: TechAmerica Reports, pp. 1–40 (2012).

12. Wang et al., , Mass Function Derivation and Combination in Multivariate Data Spaces, 8th International FLINS Conference on Computational Intelligence in Decision and Control, Madrid, Spain, 21-24 September 2008.