

A novel approach to decision making of Mined Data using Dynamic Snapshot Pattern Recognition Algorithm (DS-PRA)

Mahmoud Z. Iskandarani

Faculty of Science and Information Technology, Al-Zaytoonah University of Jordan, Amman-Jordan;

m.iskandarani@hotmail.com

ABSTRACT

A new approach to pattern recognition and decision machines profiling is proposed, proved and tested. The technique adopts the Snapshot method dynamically as a function of both organization policy and the organization department's policies. Such policies are associated with individual products and services provided by the organization with departments policies derived from general organization profile with the organization policy being a function of the various department's profiles. It is proved through real data the ability of such algorithm to classify, detect and predict policy changes and identify differences between different organizations. Also, such algorithm combines the concepts of general Artificial intelligence through the use of knowledge bases and Neural Networks by utilizing a similar weights matrix.

Keywords: Snapshot, Pattern Recognition, classification, Data Mining, Intelligent Systems.

1 INTRODUCTION

In many real-world data mining tasks, decision making may change as the time changes. This is due to change in the learned knowledge as a function of data exposure with a resulting optimization of the knowledge base (KB) as a function of experience and multi-case handling, which is a function of the everyday expansion of learned knowledge [1-3].

In order to learn and make accurate predictions, the evolution of the decision making concept must be considered, and thus, a series of data sets collected at different times is needed to keep updating the data base of learned knowledge, resulting in preserving the main characteristics of the prediction model with modifications, adjustments and optimization due to the constant input and feedback to and from the knowledge base. So a snapshot of the characteristics and features of the system under consideration would suffice to reach a decision. This would save significant time and storage and it will support the decision making process without the need to have an updated data tapped in from the knowledge base [4-6].

In this work it is shown that using the DS-PRA setup, a somewhat more natural way of temporal dependent data classification can be achieved. In addition this setup, can successfully replace all those processes of traditional classification scheme, such as, feature extraction and finding sufficient classification algorithms. It is also demonstrated here that this task can be done with a relatively small amount of data, which can be regarded as more suitable for real-time applications Moreover the model has the ability to convert a human dependent task of the selection of features into an automated computational task of finding DS-PRA parameters [7-10].

2 MATERIALS AND METHODS

The traditional methods of actual classification schemes [11-13], usually consist of the following basic parts as shown in Figure 1:

- 1) Feature Extraction
- 2) Applying classification algorithm
- 3) Decision Algorithm.

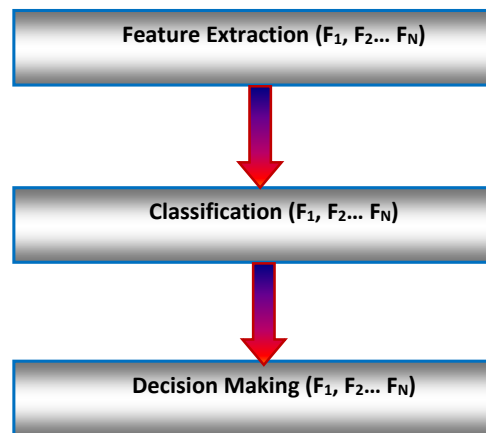


Figure 1: Traditional Decision Making System

The goal of the three steps to reduce the data explosion that is derived from real life data.

The feature extraction part requires substantial experience with the specific data task and the success of finding good features can be affected by experience, extensive analysis of the data and time-frequency manipulations. All of these are time demanding processes, and usually computationally ineffective and somewhat not natural. Their main goal is data dimensionality reduction for the classification algorithms.

The classification and decision making strictly depends on the previous one. Successful feature extraction process will result in efficient and fast learning, which affect its accuracy and generalization.

An accurate and effective model is developed from a single snapshot of data with the help of domain knowledge. The correlation of snapshots gathered over a period of time is used to establish a pattern of the organization and its decision making policy. Any change in the periodical updating of the organization decision making process will result in a change in its last snapshot and in the overall pattern and behavior of the model used.

In principle, an ensemble of data, called snapshots, collected from the database. The algorithm is then used to produce a set of basic elements that can span the original snapshot collection. It is such capability that allows extracting the representative characteristics of a decision making system of an organization. The resulting few elements that form the final profile can be regarded as dominant patterns. The proposed model and method indicate that the DS-PRA method can successfully detect pattern changes and predict decisions without the need to tap back into the database. The decision making machine profiling is affected by new data which acts as a stimulus with a dynamic continuously changing history with sliding threshold for decision making. This system is proposed and designed with an objective of achieving both system accuracy while maintaining good generalizability properties [14-18]. The overall system is illustrated in Figure 2.

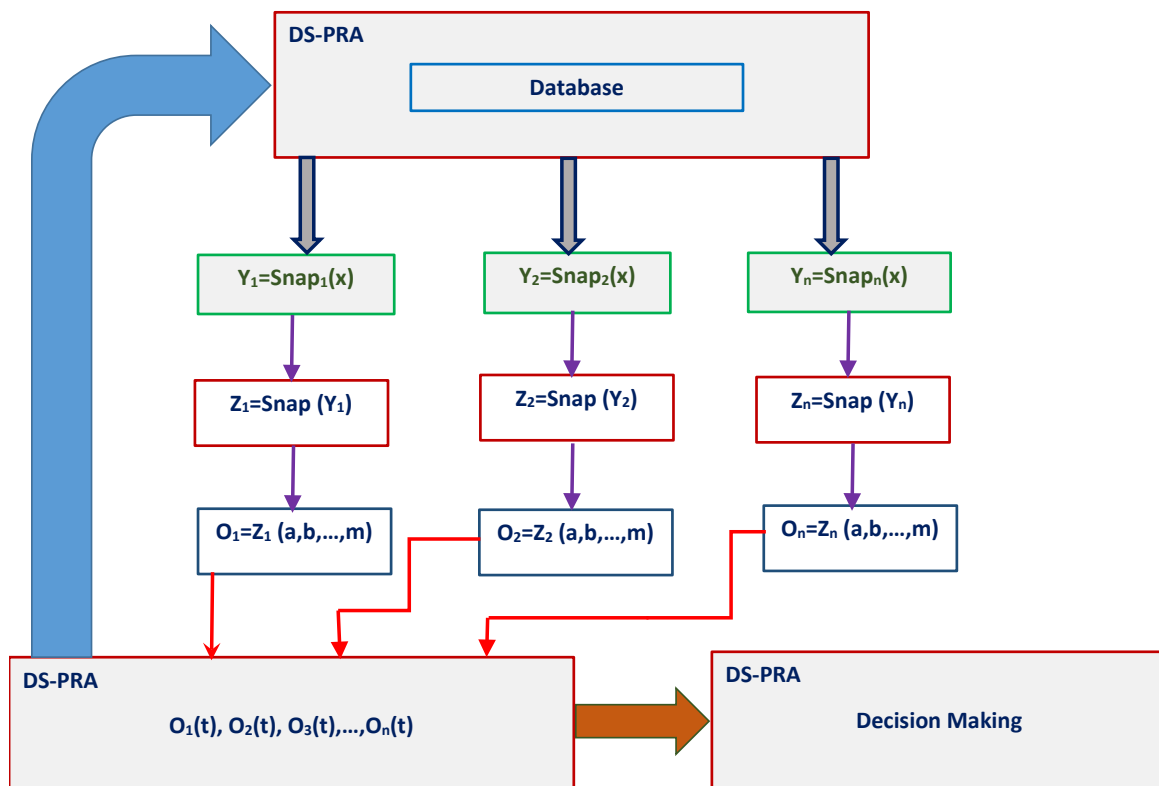


Figure 2: DS-PRA System

The resulting decisions and limits should correlate to the original policy of the organization and can be described by equation 1:

$$Decision(t_n) = f(Snapshot_{Company Policy}(t_1, t_2, \dots, t_n)) \dots (1)$$

With;

$$Snapshot_{Company Policy} = f(company Policy Variables) \dots (2)$$

For a snapshot to be valid and utilized for decision making and prediction, equation (3) must be satisfied:

$$\frac{d(Snapshot)}{dt} = 0 \dots (3)$$

The expression in (3) is used to also determine if there is a change in the decision policy if the result does not equate to zero. This indicate a change in transformation function of the concerned variables and is used to form new Snapshot if sufficient period of time is allowed to establish stability. From the previous, equation (4) is used to map the Snapshot change and to linearly predict the outcome for the new policy.

$$Var_{Policy} \left| \frac{d(Snapshot_{n+1})}{dt} - \frac{d(Snapshot_n)}{dt} \right| \leq \alpha \dots (4)$$

$$0 \leq \alpha \leq 1$$

For the prediction and decision making to be valid, the Snapshot variation for real and predicted data should not exceed the Snapshot variation limit of the general company policy and profile, hence:

$$Var_{data} \leq Var_{Policy} \leq \alpha \dots (5)$$

3 RESULTS

Tables 1 and 2 show DS-PRA Snapshot results of policies over two years for two organizations dealing with products and services, where the transactions on these products and services are affected by the organization policy that covers compensation aspects through different departments.

Each main variable have snapshots of its sub variable that contributed to its value. The snapshot of the policy is based on historical data gathering and intelligent correlation between the snapshot variables shown in tables 1 and 2.

Table 1: Original snapshots for Organization1

Organization₁		
Snapshot Variable	Policy-2012	Policy-2013
Maximum Compensation	52.43%	53.13%
Minimum Compensation	40.48%	37.48%
No Compensation	6.14%	6.65%
Rejection	0.95%	2.74%
Department₁		
Maximum Compensation	0.75%	0.14%
Minimum Compensation	88.02%	88.54%
No Compensation	4.15%	6.57%
Rejection	7.08%	4.75%
Department₂		
Maximum Compensation	86.67%	81.53%
Minimum Compensation	8.04%	12.97%
No Compensation	5.23%	4.09%
Rejection	0.06%	1.40%

Table 2: Original snapshots for Organization2

Organization₁		
Snapshot Variable	Policy-2012	Policy-2013
Maximum Compensation	46.84%	44.63%
Minimum Compensation	47.24%	43.12%
No Compensation	5.63%	10.50%
Rejection	0.29%	1.75%
Department₁		
Maximum Compensation	0.67%	0.84%
Minimum Compensation	89.60%	88.83%
No Compensation	8.72%	7.26%
Rejection	1.01%	3.07%
Department₂		
Maximum Compensation	89.73%	86.41%
Minimum Compensation	4.83%	4.80%
No Compensation	5.29%	6.95%
Rejection	0.15%	1.84%

4 ANALYSIS AND DISCUSSION

The variables are integrated into to general variables as demonstrated in tables 3 and 4. After integration, a boundary is constructed around the integrated variables to establish center points and limits. The limits are then applied to departments in each organization to enable decision making and prediction.

Table 3: Derived Snapshots for Organization1

Organization1			
Decision System	Policy Limits-Year₁	Policy Limits-Year₂	Var
Compensation	90.58%-92.91%	90.61%-92.93%	0.01%
No Compensation	9.42%-7.09%	9.39%-7.07%	
Department₁			
Compensation	88.68%-88.77%	88.68%-88.77%	0%
No Compensation	11.32%-11.23%	11.32%-11.23%	
Department₂			
Compensation	94.59%-94.71%	94.51%-94.62%	0.01%
No Compensation	5.41%-5.29%	5.49%-5.38%	

Table 4: Derived Snapshots for Organization2

Organization2			
Decision system	Policy Limits-Year₁	Policy Limits-Year₂	Var
Compensation	89.39%-94.08%	87.75%-93.17%	0.73%
No Compensation	10.61%-5.92%	12.25%-6.83%	
Department₁			
Compensation	87.59%-90.27%	89.67%-91.9%	0.45%
No Compensation	12.41%-9.73%	10.33%-8.1%	
Department₂			
Compensation	91.69%-94.56%	91.21%-94.25%	0.17%
No Compensation	8.31%-5.44%	8.79%-5.75%	

From Tables 3 and 4 and referring to equations 3, 4, and 5, the following is realized:

1. The Var for each considered department in the considered organizations is within the proposed limits and verifies the validity of the algorithm.
2. Organization₁ shows no or negligible snapshot change over two years, with one department policy suffers no change.
3. Organization₂ shows more sizable shift in the general Snapshot for the organization policy with all departments suffering Sub-Snapshot change over two years.

4. All department within both organizations have their Var_{data} that represents the departments performance in providing services Less or equal to Var_{Policy} . This satisfies the proposed algorithm and prove its validity.
5. Matrix of Sub-Snapshots is formed. This is equivalent to the matrix of weights in the Neural Network algorithms, but much simpler as shown in matrices 1 and 2 shown in expressions (6) and (7).

$$M_{Organization_1} = \begin{bmatrix} 2.33 & 2.32 \\ 0.09 & 0.09 \\ 0.12 & 0.11 \end{bmatrix} \dots(6)$$

$$M_{Organization_2} = \begin{bmatrix} 4.69 & 5.42 \\ 2.68 & 2.23 \\ 2.87 & 3.04 \end{bmatrix} \dots(7)$$

From the matrices it is realized that Organization₁ had little or no change in its services policy. This is supported by small variation per year compared to Organization₂. The weights or Sub-Snapshots are used by the algorithm to predict outcomes and make decisions dynamically as Sub-Snapshots are updated. Hence the weights or Sub-Snapshots and subsequently the matrices will be further populated and can only be limited by the choice of period of time determined by the algorithm.

The general populated and accumulated matrix for an organization is given in equation (8).

$$M_{Organization_k} = \begin{bmatrix} OrgVar_{Year_1} & OrgVar_{Year_2} & OrgVar_{Year_3} & \dots & OrgVar_{Year_n} \\ Dep1Var_{Year_1} & Dep1Var_{Year_2} & Dep1Var_{Year_3} & \dots & DepVar_{Year_n} \\ Dep2Var_{Year_1} & Dep2Var_{Year_2} & Dep2Var_{Year_3} & \dots & Dep2Var_{Year_n} \\ DepiVar_{Year_1} & DepiVar_{Year_2} & DepiVar_{Year_3} & \dots & DepiVar_{Year_n} \\ DepnVar_{Year_1} & DepnVar_{Year_2} & DepnVar_{Year_3} & \dots & DepnVar_{Year_n} \end{bmatrix} \dots(8)$$

where: Org: Organization, Dep: Department.

The Var matrix is formed as shown in equation (9) and (10).

$$Var_{Organization_1} = \begin{bmatrix} 0.00 \\ 0.01 \end{bmatrix} \leq 0.01 \leq 1 \dots(9)$$

$$Var_{Organization\ 2} = \begin{bmatrix} 0.45 \\ 0.17 \end{bmatrix} \leq 0.73 \leq 1 \dots(10)$$

The Var matrix changes over time in an accumulative way. Equation (11) describes a general form of such matrix.

$$Var_{Organization\ K} = \begin{bmatrix} Var_1 \\ Var_2 \\ Var_3 \\ Var_i \\ Var_n \end{bmatrix} \leq 1 \dots(11)$$

From the Var equations, the rest of the Var values for other departments within the organization can be predicted as it is restricted to a maximum value of 1. So, for Organization₁, it is clear that its policy is constant with hardly any variation, which leaves 0.99 span for overall policy change and for adding other departments or variables, and 0 margin of change for current Var_{Total}. For Organization₂, and as departments1 and 2 consumed most of the Var values, the span for the rest of departments is narrow and of value 0.27 to reach the maximum allowed value of 1 and 0.11 for current Var_{Total}. Such approach allows to predict and plan policies that reflect on outcomes by using parts of the organization's profile and policies and project that to obtain a dynamic curve of behavior. the algorithm is further enhanced through scheduled policy selection, hence, a variable that is selected now, will become constant next iteration and by final correlation and integration, the whole picture becomes integrated. this can be illustrated in equation (12) and (13).

$$Var_{Total} = \sum_{i=1}^n Var_i \dots(12)$$

$$Snapshot_{Total} = f(Var_{Total}) \dots(13)$$

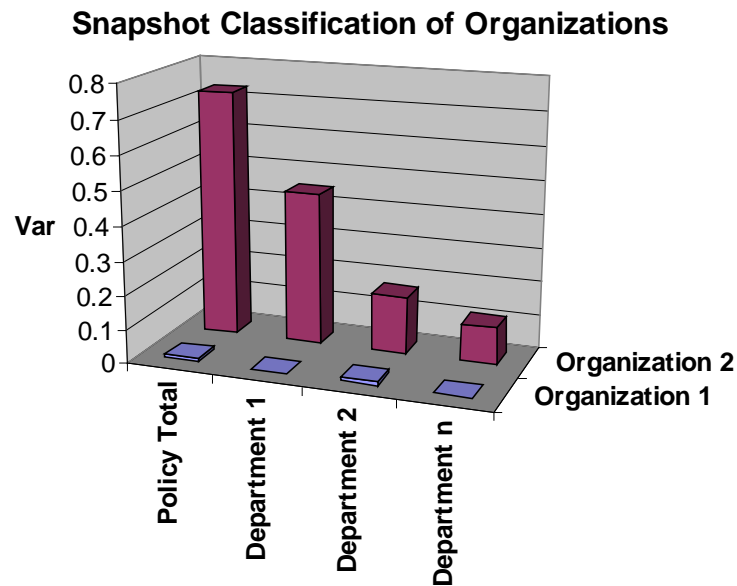


Figure 3: Classification of Two Organizations

Figure 3 shows a clear difference in two organizations policies. These organizations deals with same products but have different policy change rates. Obviously organization 1 has a constant and stable policy, which should reflect on its market dealings and its services in comparison to organization 2. such values obtained through Snapshot algorithm reflects on two main issues:

- 1) Confidence level in the organization if stability is important.
- 2) Flexibility in policy adjustment to suit market and people, if such changes are for the consumer benefits.

5 CONCLUSION

The developed algorithm proved to be effective in:

1. Reducing data size that characterizes general decision making of an organization based on historical data.
2. Enable decision making and prediction dynamically.
3. Provide and efficient method for change in decision making as a function of policy change.
4. Enable future planning of any organization policy and strategy.

REFERENCES

- [1]. J. Ramon, C. comendant, Open Problem: Learning Dynamic Network Models from a Static Snapshot. 25th Annual Conference on Learning Theory, JMLR: Workshop and Conference Proceedings, 2012. 23: p. 45.1-45.3.

- [2]. Z. Mousavinasab, H. Bahadori, AN OVERVIEW ON DATA MODELS FOR KNOWLEDGE DISCOVERY FROM DATABASES. International Journal of Advanced Research in IT and Engineering, London:2013. 2(7): p.12-20.
- [3]. G. Kou, W.Wu, TAn Analytic Hierarchy Model for Classification Algorithms Selection in Credit Risk Analysis. Mathematical Problems in Engineering, 2014. 2014(297563): p. 1-7.
- [4]. S. Beniwal, J. Arora, Classification and Feature Selection Techniques in Data Mining. International Journal of Engineering Research & Technology (IJERT), 2012. 1(6): p. 1-6.
- [5]. A. Moeinian, S. Baladehi, A. Zolfagharian, Hybrid Genetic Algorithm Using the Solving Open Shop Scheduling. International Journal of Engineering Research and Technology, 2013. 5(2): p. 1-10.
- [6]. V. Vasani, R. Gawali, Classification and performance evaluation using data mining algorithms. International Journal of Innovative Research in Science, Engineering and Technology, 2014. 3(3): p. 10453-10458.
- [7]. R. Kumar, R. Verma, Classification Algorithms for Data Mining:A Survey. International Journal of Innovations in Engineering and Technology (IJET), 2012, 1(2): p. 7-14.
- [8]. R. Lokeshkumar, P. Sengottuvelan, M. Vina, An Approach for Web Personalization using Relational Based Fuzzy Clustering Ontology Model, Biomedical Engineering, Australian Journal of Basic and Applied Sciences, 2014. 8(2): p. 18-22.
- [9]. M. Peker, B. Sen, S. Bayir. Using Artificial Intelligence Techniques for Large Scale SetPartitioning Problems. Procedia Technology, 2012 , 1: p. 44 – 49.
- [10]. N. Davuth, K. Sung-Ryul, Classification of Malicious Domain Names using Support Vector Machine and Bi-gram Method, International Journal of Security and Its Applications, 2013, 7(1): p. 51-58.
- [11]. H. Yang, S. Fong, G. Sun, R. Wong, A Very Fast Decision Tree Algorithm for Real-Time Data Mining of Imperfect Data Streams in a DistributedWireless Sensor Network. International Journal of Distributed Sensor Networks, 2012, 2012(863545): p.1-16.
- [12]. S. Dandu, B. Deekshatulu, P. Chandra, Improved Algorithm for Frequent Item sets Mining Based on Apriori and FP-TreeIEEE. Global Journal of Computer Science and Technology Software & Data Engineering, 2013. 13(2): p. 1-5.
- [13]. Y. Yang, Z. Zhi-Hua, A Framework for Modeling Positive Class Expansion with Single Snapshot. Springer-Verlag Berlin Heidelberg, 2008, p. 429-440.
- [14]. L. Zhenliang, B. Wang, X. Xiaowei, P. Hannam, Environmental emergency decision support system based on Artificial Neural Network. Safety Science, 2012. 50(2012): p. 150-163.
- [15]. S. Gupta, D. Kumar, A. Sharma, DATA MINING CLASSIFICATION TECHNIQUES APPLIED FOR BREAST CANCER DIAGNOSIS AND PROGNOSIS. Indian Journal of Computer Science and Engineering (IJCSE), 2011. 2(2): p. 188-195.

- [16]. X. Wu, et al., Top 10 algorithms in data mining. Knowl Inf Syst, 2008, 14: p.1–37.
- [17]. S. Strohmeier, F. Piazza, Domain driven data mining in human resource management: A review of current research, Expert Systems with Applications, 2013, 40(7):p.2410-2420.
- [18]. H. Tsai, Knowledge management vs. data mining: Research trend, forecast and citation approach. Expert Systems with Applications, 2013, 40(8): p. 3160-3173.