Machine Learning Approach for Process Modeling

P.V.Kumaraguru,
Research Scholar,
Dr.M.G.R. Educational and Research Institute University
Chennai-600095. India.

S.P.Rajagopalan, PhD.
Professor Emeritus,
Dr. M.G.R. Educational and Research Institute University
Chennai- 600095. India.

ABSTRACT

All the modern business industries are in the urge to find the ways and means to handle the digital data which are generated automatically for every business transaction. These transactional data are recorded in the temporal database as event logs along with a unique case and event ID for future reference. Every process will leave its foot prints in the event logs, the size of these logs has grown unimaginable huge and triggered many challenges for the researchers. Providing only hardware solutions for storage is not wise enough, instead extracting process models and enhancing it through machine learning techniques is the real challenge of the day. This paper has made an attempt to find the suitable notations for modeling the process and suggests various notations for process model which intern leads for process optimization.

Keywords

Business intelligence, Temporal Database, Data explosion, Event logs, process model, machine learning, stochastic.

1. INTRODUCTION

Process optimization involves the clear understanding of the flow and functioning of the business process, and selecting a suitable notation for transforming the business process in to an unanimous visual model. This model should make use of limited standard symbols that should be able to understand and analyze it. By reorganize or reengineering the process model the throughput of the process can be enhanced. Process optimization will lead to process evaluation.

2. MACHINE LEARNING

Machine learning is a field of study that gives computers the ability to learn without being explicitly programmed. Machine learning has been classified in to supervised learning and unsupervised learning. Supervised learning uses trained set of data where as unsupervised learning uses unlabeled data set. Regression which uses continuous values is a best method for supervised learning. Clustering is suitable for unsupervised learning.

3. PROCESS MODEL NOTATIONS

3.1 Markov Chain

Markov Chain is a new type of change process, the outcome of a given experiment can affect the outcome of the next experiment, It is a special kind of stochastic process, which evolve over time in a probabilistic manner. In Markov chain model the outcome of an experiment depends only on the outcome of a previous experiment

3.2 Petri Nets

Petri Nets are graphical and mathematical modeling notations.

A Petri net is a four-tuple:

$PN = <P, T, I, O>$

P: a finite set of places, $\{p1, p2, ...,pn\}$

T: a finite set of transitions, $\{t1, t2, ...,ts\}$

I: an input function, $(T \times P) \rightarrow \{0, 1\}$

O: an output function, $(T \times P) \rightarrow \{0, 1\}$

Petri nets are traditionally use for describing and analyzing systems that are characterized as concurrent, asynchronous, distributed, parallel nondeterministic and/or stochastic. Due to their graphical nature, Petri Nets can be used as a visualization technique like flow charts or block diagrams but with much more scope on concurrency aspects. As a strict mathematical notation, it is possible to apply formal concepts like linear algebraic equations or probability theory for investigating the behavior model led system. In the last decade, Petri Nets have become a powerful concept in the area of business process modeling and work flow management.

Petri nets have been successfully used to model and analyze processes from many domains such as software and business processes especially work flow processes.

Petri Net model consist of two parts

(1) The net structure that represents the static part of the system and

(2) A marking that represents the overall state on the structure. The token distribution among the places of a Petri net is called its marking. When one or more tokens reside in a place, the place is said to be marked, otherwise it is unmarked.

Petri nets have been successfully used to model and analyze processes from many domains such as software and business processes especially work flow processes.

The classical Petri net is a directed bipartite graph with two node types called *places* and *transitions*. The nodes are connected via directed *arcs*. Connections between two nodes of the same type are not allowed. Places are represented by circles and transitions by rectangleS.

4. PETRINET NOTATIONS

A Petri Net consists of places, transitions and arcs that connect them.

Places: Places are represented by ellipses (often circles).Three annotations are associated with a place p:

the place name;the name of the type (Type(p)) associated with the place; and

the initial marking.

Transitions: Transitions are rectangles and arcs as arrows. Input arcs connect transitions with places.

Arcs: An arc is represented by an arrow.

Places are passive components and are modeling the system state. They can contain TOKENS, depicted as black dots. The current state of the Petri net is given by the number of tokens on each place. Transitions are active components modeling activities which can occur and cause a change of the state by a new assignment of token to places. Transitions are only allowed to occur if they are enable, which means that there is at least one token on each input place. By occurring, the transition removes a token from each input place and adds a token on each output place. The repeated occurrence of transitions and the resulting sequence of marking are called the token game.

Petri Net model consist of two parts

(1) The net structure that represents the static part of the system and

(2) A marking that represents the overall state on the structure. The token distribution among the places of a Petri net is called its marking. When one or more tokens reside in a place, the place is said to be marked, otherwise it is unmarked.

Transitions symbolize actions; places symbolize states or conditions that need to be met before an action can be carried out

5. TEMPORAL DATABASE OF HEALTH INSURANCE CLAIM

To illustrate the graphical representation of the trace table we consider the health insurance claim processing activity in an insurance company.

To provide prompt claims servicing to the insures, the insurance companies appoint the Third Party administrator (TPA) duly licensed by the Insurance regulatory and development authority (IRDA) of India or with the help of the Health Claims Management Team (HCMT). Insurance company provides hassle free claims services to its customers through cashless facility.Cashless facility is the service wherein the customer need not pay the hospitalization bills either at the time of admission or at the time of discharge from the hospital. This facility can be availed only from its network hospitals

Initiallly the insured intimates the TPA / HCMT (task *Register the claim request with TPA / HCMT*)

The trace table is derived from a fragment of insurance event log which has timestamp of various activities. In the insurance event log every case has a unique id and every has several event. Every event has also have a unique event id. The sample database has eight distinct activities refered through a to h.

CASE ID	TRACE
1	(a,b,c,d,e)
2	(a,c,b,d,f)
3	(a,b,c,d,g,h,i,d,j)
4	(a,b,c,d,g,i,h,d,j)
5	(a,b,c,d,g,i,h,d,k)

The various labels mentioned in the table denotes as

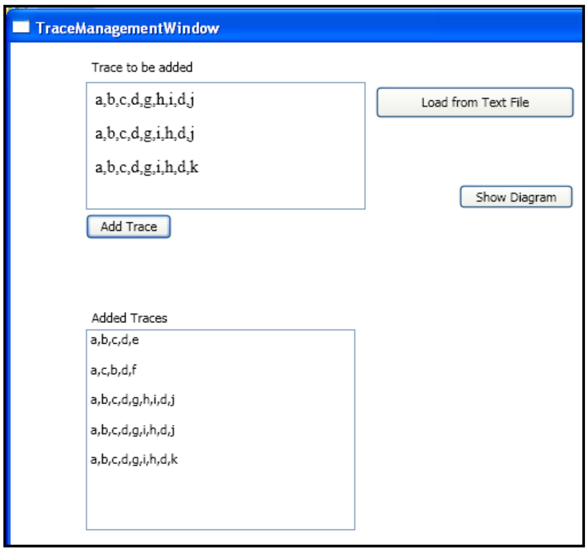
- a -Register the claim request with TPA / HCMT.
- b -Examine the policy number and the nature of sickness.
- c -Check if the sickness is insured under the policy.
- d - Decide.
- e -Repudiate the claim.
- f -Honor the claim.
- g -Check if it is a network hospital.
- h -Examine.
- i-Verify the Cashless Facility Form.
- j - Honor the cashless facility.
- k - Repudiate the cashless facility.

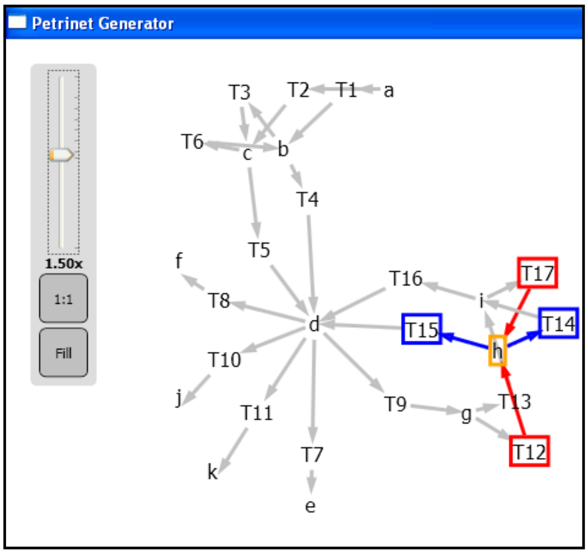
Table : A fragment of Health Insurance Claim Process -each line corresponds to an event.

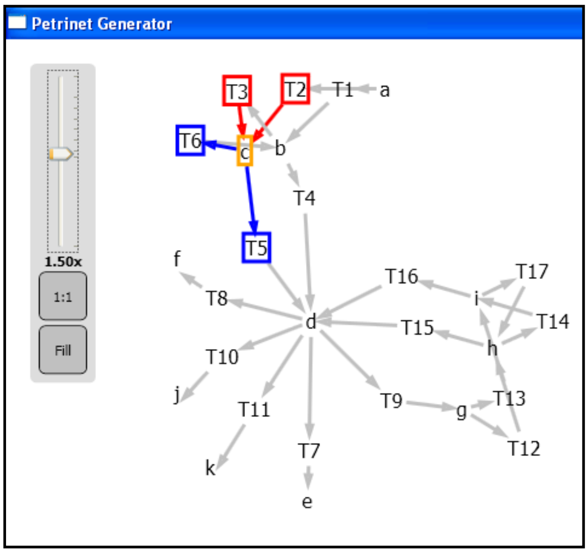
CASE ID	EVENT ID	DATE	TIME	ACTIVITY	RESOURCE
1	5634550	20/8/2012	11:05	Register the claim request with TPA / HCMT	Raj
	5634551	20/8/2012	11:30	Examine the policy number and the nature of sickness	Mano
	5634552	20/8/2012	11:45	Check if the sickness is insured under the policy	Sam
	5634553	20/8/2012	12:10	Decide	Ravi
	5634554	20/8/2012	12:30	Repudiate the claim	Issac
2	5634600	17/9/2012	9:10	Register the claim request with TPA / HCMT	Ravi
	5634601	17/9/2012	9:30	Check if the sickness is insured under the policy	Raj

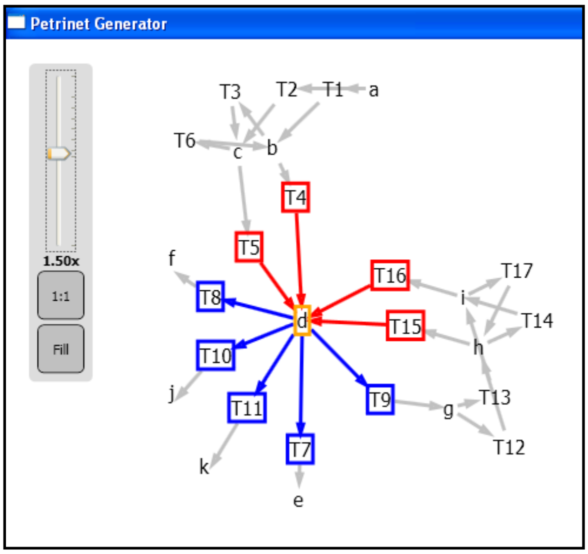
	5634602	17/9/2012	9:45	Examine the policy number and the nature of sickness	Sam
	5634603	17/9/2012	10:30	Decide	Mano
	5634604	17/9/2012	11:00	Honor the claim	Mani
	5634321	12/9/2012	15:00	Register the claim request with TPA / HCMT	Ravi
	5634322	12/9/2012	15:10	Examine the policy number and the nature of sickness	Mani
	5634323	12/9/2012	15:30	Check if the sickness is insured under the policy	Sam
	5634324	12/9/2012	15:50	Decide	Mano
3	5634325	12/9/2012	16:30	Check if it is a network hospital	Issac
	5634326	12/9/2012	16:45	Examine	Mano
	5634327	12/9/2012	17:45	Verify the Cashless Facility Form	Ravi
	5634328	12/9/2012	18:10	Decide	Mani
	5634329	12/9/2012	18:45	Honor the cashless facility	Issac
	5634775	6/7/2012	11:15	Register the claim request with TPA / HCMT	Mano
	5634776	6/7/2012	11:35	Examine the policy number and the nature of sickness	Mani
	5634777	6/7/2012	12:10	Check if the sickness is insured under the policy	Issac
	5634778	6/7/2012	12:30	Decide	Raj
4	5634779	6/7/2012	13:00	Check if it is a network hospital	Ravi
	5634780	6/7/2012	13:25	Verify the Cashless Facility Form	Raj
	5634781	6/7/2012	13:45	Examine	Ravi
	5634782	6/7/2012	14:15	Decide	Mano
	5634783	6/7/2012	14:45	Honor the cashless facility	Mano
	5634990	30/09/2012	17:25	Register the claim request with TPA / HCMT	Issac
	5634991	30/09/2012	18:10	Examine the policy number and the nature of sickness	Raji
	5634992	30/09/2012	18:30	Check if the sickness is insured under the policy	Ravi
	5634993	30/09/2012	18:45	Decide	Mano
5	5634994	30/09/2012	19:15	Check if it is a network hospital	Mano
	5634995	30/09/2012	19:40	Verify the Cashless Facility Form	Issac
	5634996	30/09/2012	20:00	Examine	Mani
	5634997	30/09/2012	20:15	Decide	Mani
	5634998	30/09/2012	20:40	Repudiate the cashless facility	Raji

The trace table is graphically represented as below









6. CONCLUSION

In this paper we discussed the about types of machine learning and the process model notations particularly Markov chain and Petri nets. Petri nets are most suitable for machine learning approach because of its properties such as concurrent, asynchronous, distributed, parallel nondeterministic and stochastic. These process models can be further tuned using machine learning techniques such as unsupervised learning. Hence the process can be re engineered and process perfection can be archived. Cost and time effective process can be modeled through machine learning which does not expect explicit programming.

7. REFERENCES

[1] Machine learning – John Anderson.

[2] Process modeling from insurance event log – International Journal of Computer Applications.

[3] Wil M.P. Van der Aalst, 2010,Process Mining, Springer.

[4] W.M.P. Van der Aalst 1998 The applications of Petri Nets to wotk flow management, The journal of Circuits, Systems and computers.

[5] W.M.P. Van der Aalst and K.M.VanHee. Workflow Management: Models, Methods, and system. MIT press, Cambridge, MA, 2002.

[6] S.Kumaran and K.Raja, Modeling and simulation of Projects with Petri Nets. American journal of applied Sciences, 2008.

[7] A.J.M.MWeijters and W.M.P Van der Aalst, Process Mining : 2001 Discovering workflow Models from Event based Data, 13th Belgium- Netherlands Conference on Artificial intelligence(BNAIC 2001).

[8] W.M.P van der Aalst A.J.M.M. Weijters, and L.Maruster. 2004, Work flow Mining: Discovering process models from Event Logs. IEEE Transactions on Knowledge and Data Engineering.