21/5/2024

Process Mining

Type of events

i) casual: (a & b are events)
if a comes before b then b can not come before a. (a → b)

ii) parallel:
can happen simultaneously

iii) choice:
a & b arenot consecutive

Steps:
i) define all events
ii) Identify start event
iii) identify all possible end events
iv) calculate footprint table
v) make sets
vi) Drop max. sets
vii) Draw petrinets

event s:

<A, B, C, D>

<A, C, B, D>

<A, B, D>

<A, C, D>

A → B , A → C , B || C

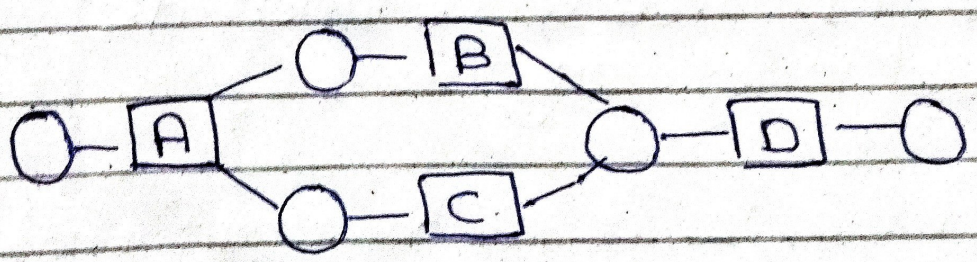
C → D , B → D

	A	B	C	D
A	#	→	→	#
B	←	#	||	→
C	←	||	#	→
D	#	←	←	#

Sets:

(A, B) , (A, C) , (B, D) , (C, D)

Drop Sets



Data

- Data
  - collection of facts & figures that represent something in real world
  - can be numbers, symbol, text, images.

- Datapoint:
  - individual unit or piece of data.
  - can be simple number (temprature) or complex (sentence in document)

- Importance of structure & organization
  - Data exists in raw form (collection of datapoint)
  - it gains its true value when it's structured and organized

- Unstructed: doesnot follow any format.
- Semistructured: has some internal (organization) format but doesnot conform to strict format

Sources of data:

- events over the Internet (Internet of events).

Two Type

↓

Data Centric
(focuses an amount of exact data generated from source)

- Internet of Content
- Internet of people
- Internet of things
- Internet of location

↓

Process Centric
(use of data to analyze & improve process)

1). IoC - all information created by humans to increase knowledge i.e Wikipedia, youtube, ebook etc.

2). IoP - all data related to social interaction i.e email, facebook etc

3) IoT - all physical objects connected over internet

4) IoL - all data that have geographical & spatial dimension.

so IOE in this context means, collecting, analyzing, utilizing data to gain insights, understand process and make decisions

Data Centric
- all event data

Process Centric
- analyzing event sequence.

Big data

- refers to complex, voluminous & rapidly growing that traditional processing techniques are inadequate for it.

- It is characterized by 3V's:
  i). Volume (massive data)
  ii) Variety (Structured, semi Structured, UnStructured)
  iii). Velocity (speed at which data is being generated).

Data Science

- interdisciplinary field aiming to convert data into real values
- value can be in form of predictions, automated decisions, models learned from data.

- includes:
  - data extraction
  - data preparation
  - exploration
  - transformation

- storage & retrival
- computing infrastructure
- various type of mining & learning
- presentation of explaination & prediction

Data mining - process of extracting patterns or insights from large dataset using some algorithm.

Process Mining - Missing link b/w data mining / data science & process science

Process Science - focuses on understanding, designing and improving business process

Simulations (virtual models)

- imitation of operation of real world process.

Purpose (Why we do it):
i). to study behaviour of system
ii). to study behaviour & operation without impacting real world
iii). allows experimentation.
iv). can be done at low cost using a software

Purpose of studying behaviour:
- allows studying of complex system and modeling & understanding systems that are difficult to study in real world due to their intricate interations & variables.
- helps in understanding various conditions and making informed decisions
- insights are used for improvements.

Importance:
- understand complex systems
- reduce risk
- cost efficiency
- time saving
- training and education
- Testing, Optimization and Innovations

- Scenario analysis

Other ways of studying behaviour:
i) BPMN models
ii). flowcharts, Data flow diagrams.
but all of these are symbolic (graphs, notations. cannot capture dynamic behaviour)

Analytical Model (i.e Queue models)
- use mathematical equations

Simulation (combination of both symbolic & analytical model)

System (group of objects joined together for accomplishment of a task)
↳ Components:
i). Entities (real world objects)
ii). Attributes (properties of system)
iii). Activities (Something happening over period of time)
iv). Events (anything that brings a change in system)
v). State (description of entities, attributes, activities at one point in time)

Types of Event Simulations

i). Event simulation:
- Discrete Event Simulation
  - process of codifying a complex system as an ordered set of well defined events
  - OR how a system with discrete flow units or jobs evolves over time.

- State variables change as a result of event.

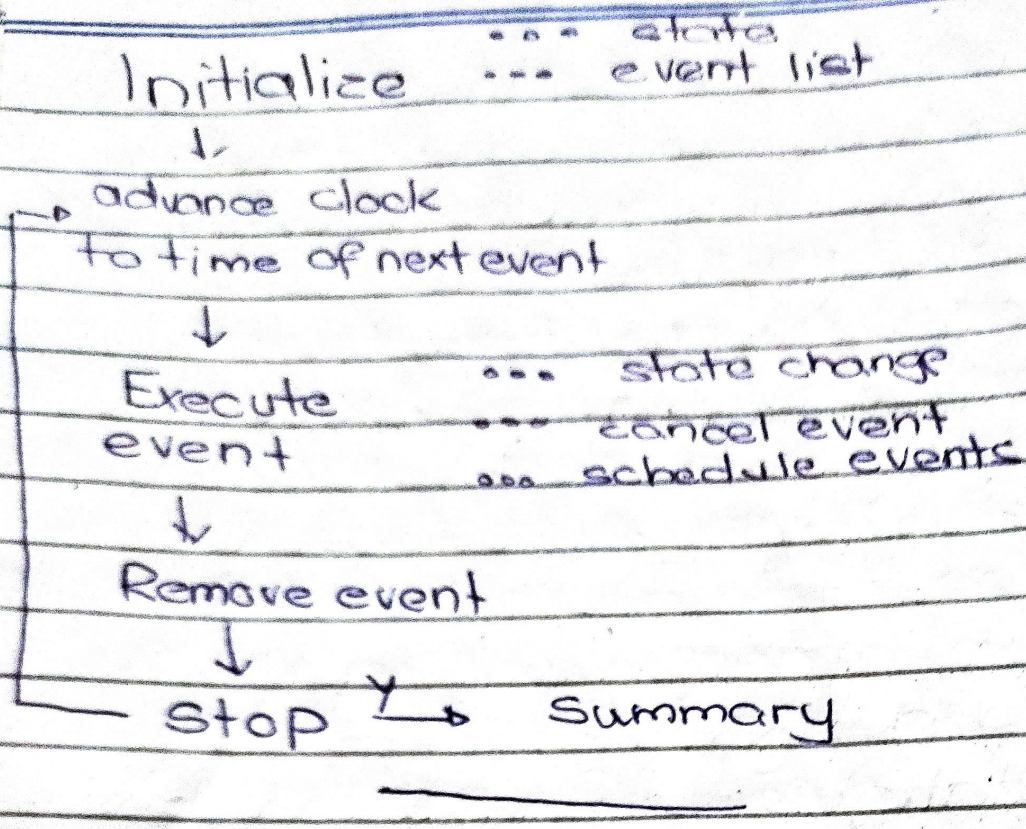
- State variables in Queueing Model:
  - Arrival rate, service rate, queue length

- In DES, we can keep a track of how and when state variables change (they usually change a result of an event).
- DES is not concerned about how much time customer spends in line, service, entire system
- only concerned about time when event occurs.

- so as it is not concerned about any other time it assumes that nothing has happened between two consecutive events.
- DES can also be called Event-Driven Simulation (Why? because it considers state variables, and state variables change when event(s) occur)

DES is designed on components:

- System State - collection of variables that represent state of system
- Simulation clock - tracks time in simulated system
- Event list - chronological (ordered) list of events scheduled to occur in simulation
- Initialization routine - defines how first event (at time 0) is selected.
- Report generator
- Random number generator (optional)
- Main program - invokes the model's routine to perform simulation.



Monte Carlos Simulation

- process of playout scenarios (hundred, thousands or millions of time)
- to get uncertainity / variability in inputs to reach closer to reality
- done using random numbers.
- has three cases
  i). Best case (lowest)
  ii). Most likely (estimated)
  iii). Worst (highest)

steps of Monte-Carlos:

1). Establish Probability distribution
2). Cumulative Probability distribution
3). Setting Random Number Intervals
4). generating Random Numbers
5). finding answer.

Pseudocode for Monte Carlos

1). Define parameters (best case (lowest value), most likely, worst case)
2). Initialize (set) number of simulations
3) generate random numbers:
for i (1 to number of simulations)
3). Calculate mean of sample and standard deviation.

Simulation: Designing

1). Define Problem (what needs to be solved)
2). ↳ list mod questions model will answer
↳ Questions are either key (must be answered) or desirable
2). Understand process
↳ identify entities
↳ define/design simulation only to answer question instead of imitating the whole system
↳ Use Pareto's law:
80% of behaviour can be understood by action of 20% of components

3). Determine goal & objectives:
↳ define evaluation criteria
↳ key performance measures

4) Obtaining support from management
- Educate management
- set realistic expectation
- highlight similar projects
- communicate

5). Choosing Simulation Software

6). Data
- Data Sources
  - i). Direct observation
  - 2). time studies & history
  - 3) employee estimates (often unreliable)
  - 4). vendor claims (overly optimistic abt reliability)
- Challenges with Data:
  - 1) limited availability
  - 2). Data Quality
  - 3) Data source reliability

Incase of Insufficient Data:

7). Assumptions

8). Determining desired output

9). Building Model

10). Project Kickoff.