Artificial Intelligence
Assignment - 3 - BAYESIAN NETWORKS

Problem 3 has been chosen.

Description:

You are the TA of Artificial Intelligence and you have prepared an assignment on Bayesian networks. You want the hard deadline of assignment to be on April 11th. But however, you know that the students of UG2k17 will ask for an extension, in view of this you decide to keep the assignment deadline ‘x’ days(x=1/2/3) prior to April 11th. Your job is to model a bayesian network to decide ‘x’ considering factors like other assignment deadlines, student’s work load, exams/quizzes schedules, fests/events, whether the topic was taught in class etc. Give conditional probabilities and justify. (We TAs have solved this problem and found out that P(x=0) is higher
given the conditions. Hence this is a hard deadline :D))

Solution:

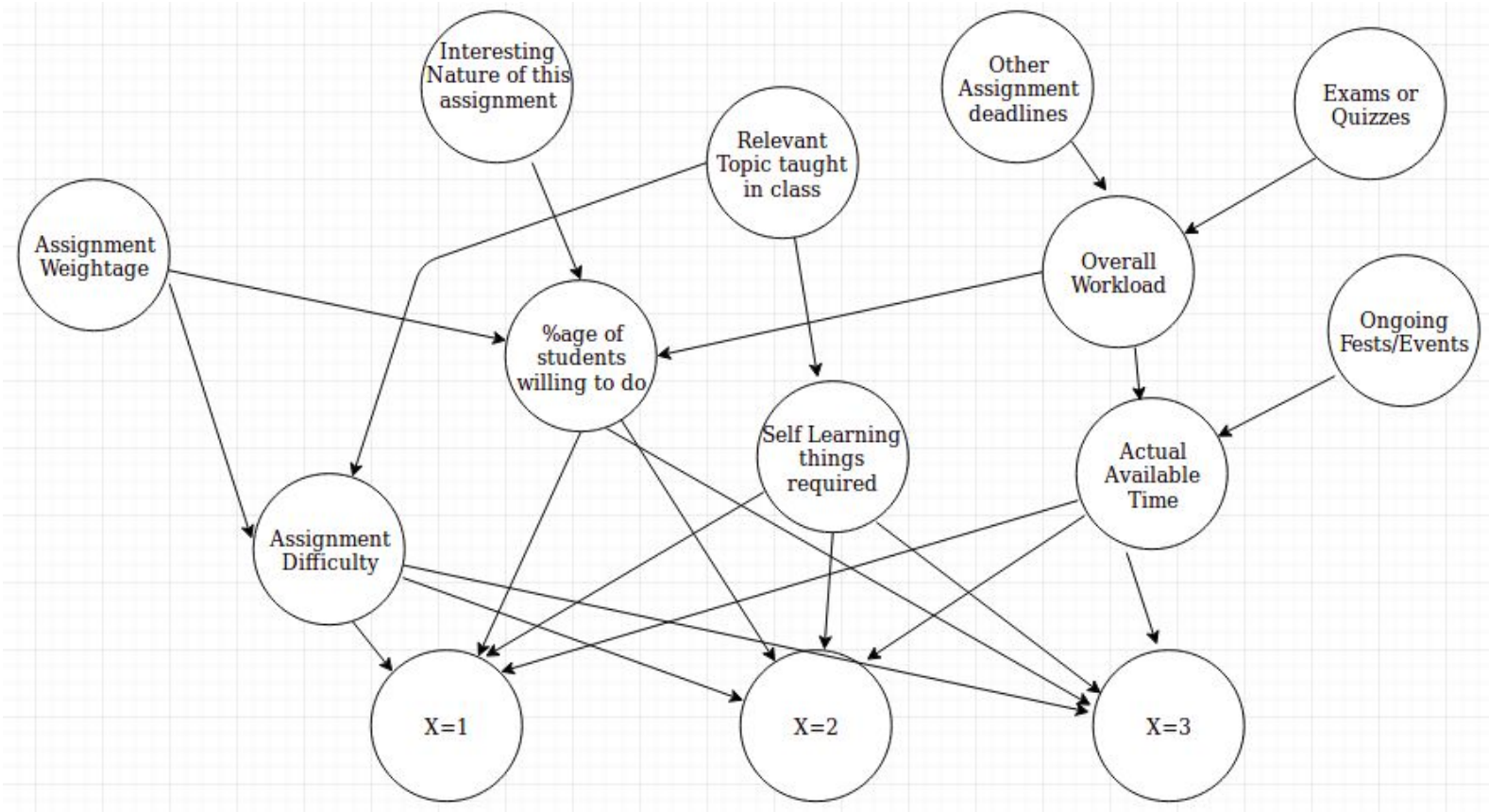

Diagram of the Bayes’ Net we have developed.

KEY:

AW	Assignment Weightage	Low, Medium, High
INT	Interesting Nature of Assignment	High, Low
RTT	Relevant Topic Taught in class or not	Yes, No
OAD	Other Assignment Deadlines	Too many, Okayish
EQ	Exams or Quizzes	Yes, No
FE	Ongoing Engaging Fests/Events	Yes, No
AD	Assignment Difficulty	Low, High
PSW	Percentage Students Willing to do	High, Low
SL	Self-Learning New Things Involved	Too much, Manageable, Very Little
OW	Overall Workload	High, Medium, Low
AT	Actual Time available after assignment release	Too much, Satisfactory, Very Little
X=1	Assignment Deadline 1 day before 11th April	Yes, No
X=2	Assignment Deadline 2 days before 11th April	Yes, No
X=3	Assignment Deadline 3 days before 11th April	Yes, No

Conditional Probability Tables:

Assignment Weightage:

AW	Low	Medium	High
P(AW)	0.2	0.4	0.4

Interesting nature of Assignment:

INT	High	Low
P(INT)	0.6	0.4

Relevant Topic Taught in class :

RTT	Yes	No
P(RTT)	0.7	0.3

Other assignment deadlines:

OAD	Too Many	Okayish
P(OAD)	0.55	0.45

Exams or quizzes:

EQ	Yes	No
P(EQ)	0.2	0.8

Ongoing Engaging Fests/Events:

FE	Yes	No
P(FE)	0.4	0.6

Assignment Difficulty:

AW	RTT	AD.High
High	Yes	0.7
High	No	0.9
Medium	Yes	0.6
Medium	No	0.7
Low	Yes	0.3
Low	No	0.5

Overall Workload:

OAD	EQ	OW.High	OW.Medium	OW.Low
Too many	Yes	0.8	0.1	0.2
Too many	No	0.6	0.25	0.15
Okay	Yes	0.6	0.25	0.15
Okay	No	0.2	0.35	0.45

Self learning things required

RTT	SL.Too much	SL.Manageable	SL.Very Little
Yes	0.2	0.6	0.2
No	0.6	0.3	0.1

Percentage of Students Willing to do - PSW:

AW	INT	OW	PSW.high	PSW.Low
High	High	High	0.7	0.3

High	High	Medium	0.75	0.25
High	High	Low	0.9	0.1
High	Low	High	0.6	0.4
High	Low	Medium	0.9	0.1
High	Low	Low	0.99	0.01
Medium	High	High	0.5	0.5
Medium	High	Medium	0.8	0.2
Medium	High	Low	0.95	0.05
Medium	Low	High	0.5	0.5
Medium	Low	Medium	0.6	0.4
Medium	Low	Low	0.65	0.35
Low	High	High	0.55	0.45
Low	High	Medium	0.68	0.32
Low	High	Low	0.75	0.25
Low	Low	High	0.1	0.9
Low	Low	Medium	0.25	0.75
Low	Low	Low	0.4	0.6

X=1:

AD	PSW	SL	AT	X1
-	Low	-	-	1
High	High	Too much	Too much	0.1
High	High	Too much	Satisfactory	0.1
High	High	Too much	Very Little	0.8
High	High	Manageable	Too much	0.1
High	High	Manageable	Satisfactory	0.1
High	High	Manageable	Very Little	0.5
High	High	Very Little	Too much	0.7

High	High	Very Little	Satisfactory	0.25
High	High	Very Little	Very Little	0.8
Low	High	Too much	Too much	0.05
Low	High	Too much	Satisfactory	0.05
Low	High	Too much	Very Little	0.9
Low	High	Manageable	Too much	0.1
Low	High	Manageable	Satisfactory	0.2
Low	High	Manageable	Very Little	0.25
Low	High	Very Little	Too much	0.7
Low	High	Very Little	Satisfactory	0.8
Low	High	Very Little	Very Little	0.9

X=2:

AD	PSW	SL	AT	X2
High	High	Too much	Too much	0.1
High	High	Too much	Satisfactory	0.65
High	High	Too much	Very Little	0.1
High	High	Manageable	Too much	0.25
High	High	Manageable	Satisfactory	0.65
High	High	Manageable	Very Little	0.3
High	High	Very Little	Too much	0.15
High	High	Very Little	Satisfactory	0.5
High	High	Very Little	Very Little	0.15
Low	High	Too much	Too much	0.9
Low	High	Too much	Satisfactory	0.8
Low	High	Too much	Very Little	0.04
Low	High	Manageable	Too much	0.15
Low	High	Manageable	Satisfactory	0.45
Low	High	Manageable	Very Little	0.25
Low	High	Very Little	Too much	0.2
Low	High	Very Little	Satisfactory	0.1
Low	High	Very Little	Very Little	0.03

X=3:

AD	PSW	SL	AT	X3
High	High	Too much	Too much	0.8
High	High	Too much	Satisfactory	0.25
High	High	Too much	Very Little	0.1
High	High	Manageable	Too much	0.65
High	High	Manageable	Satisfactory	0.25
High	High	Manageable	Very Little	0.2
High	High	Very Little	Too much	0.15
High	High	Very Little	Satisfactory	0.25
High	High	Very Little	Very Little	0.05
Low	High	Too much	Too much	0.05
Low	High	Too much	Satisfactory	0.1
Low	High	Too much	Very Little	0.01
Low	High	Manageable	Too much	0.75
Low	High	Manageable	Satisfactory	0.35
Low	High	Manageable	Very Little	0.5
Low	High	Very Little	Too much	0.1
Low	High	Very Little	Satisfactory	0.1
Low	High	Very Little	Very Little	0.02

Justifications:

1. The problem statement was quite open ended - so we considered the above mentioned factors to be affecting the choice of a particular X while fixing the deadline for assignment. Hence all the initial tables have been filled according to our general understanding and experience.
2. For the other assignment deadlines (OAD) - we have only two options - Too many and Okay and not a third option - because we always have one or more assignments going on.
3. The amount of self learning required depends to a great extent on whether the related relevant topic was taught in the class or not.
4. The actual time which a student can spend a given assignment depends not only on the overall workload on the student (due to other subjects and/or exams quizzes), but also

on whether some event is happening or not in the college - which student might be very interested to take part in.

5. To decide for final layer - consisting of different X values, following is our logic:
    a. More is the percentage of students willing to do the assignment seriously, greater will be the value of X to be set by TAs - because they want the students to start working early on it - as they know anyways that students are going to ask for a deadline extension. If this PSW is low (20-30% which is rare)- the TAs simply keep the nearest deadline (X=1 to be best) to hard deadline because they don't have to worry about too much of extension requests.
    b. TAs do not want students to simply keep complaining about short time given to them given the workload - actual time available. TAs will make the deadline appear more realistic to the students - this will take into account the amount of hard work required as well as available time:
        i. If workload will be high but manageable then X will be higher value to push students to work fast, having a chance to increase deadline will enable TAs to make students happier and motivate them to complete on time.
        ii. If less time is available for a given workload genuinely, then the TAs won't take a risk, and be more realistic with students (so that they don't simply leave the assignment) and then value of X would be kept low, to still make deadline feasible to the students.
        iii. And if the workload presented to students would be low, TAs again won't think too much, and try to keep a lower value of X - as the students need not be motivated so much then, and they can manage on their own, and complete things on time.
6. Using all these pruning methods - we reduced number of rows in our CPT table from 54 to 19.
7. The sum of probabilities of rows across all values of X sums to 1.
8. The assignment nature and difficulty level probabilities have been set according to the normal IIIT standards of the assignments in general.

Query Structure:

We have picked the following query structure to demonstrate a sample query for the given bayesian network:

P(X | p(X), p(p(X))) where p(X) refers to the parent of X.

Here X = 2 (for example)

P(X=2 | AT = Satisfactory, OW = Medium)

Next few pages will show the query solved in detail.

QUERY = P(X=2 | AT = Satisfactory, OW = medium)

Table for X=2

AD	PSW	SL	AT	P(X2=YES)
High	High	Too much	Satisfactory	0.65
High	High	Manageable	Satisfactory	0.65
High	High	Very Little	Satisfactory	0.5
Low	High	Too much	Satisfactory	0.8
Low	High	Manageable	Satisfactory	0.45
Low	High	Very Little	Satisfactory	0.1

Each row of the above table is calculated and added to get the answer

For each row in this table, we will calculate required probabilites as follows;-

P(YES) * 1 * P(SL = Too much/Manageable/Very Little) * P(PSW = High | OW = Medium) * P(AD = high/low)

P(AT=Satisfactory) = 1 (because of our assumption)

Calculating P(SL = Too Much)

RTT	SL.Too much		RTT	SL.Too much	
Yes	0.2	becomes	0.7	0.2	0.14
No	0.6		0.3	0.6	0.18
					0.32

Calculating P(SL = Manageable)

RTT	SL.Manageable		RTT	SL.Manageable	
Yes	0.6	becomes	0.7	0.6	0.42
No	0.3		0.3	0.3	0.09
					0.51

Hence P(SL = Very Little) = 1-P(SL = Too Much) - P(SL = Manageable) = 0.17

Calculating P(PSW = High | OW = medium)

AW	INT	OW	PSW.high
High	High	High	0.7
High	Low	High	0.6
Medium	High	High	0.5
Medium	Low	High	0.5

Low	High	High	0.55
Low	Low	High	0.1

Now we need to find P(OW = High)

OAD	EQ	OW.High		OAD	EQ	OW.High	
Too many	Yes	0.8		0.55	0.2	0.8	0.44
Too many	No	0.6	which becomes	0.55	0.8	0.6	0.264
Okay	Yes	0.6		0.45	0.2	0.6	0.054
Okay	No	0.2		0.45	0.8	0.2	0.072
							0.83

Hence required probability from each row = P(OW.high) * P(EQ) * P(OAD)

Hence P(OW = High) = 0.83

Now to calculate P(PSW=High | OW = medium) we can substitute values into the table of PSW

Calculating P(PSW = High | OW = medium) - with values filled

AW	INT	OW	PSW.high	
0.4	0.6	0.83	0.7	0.13944
0.4	0.4	0.83	0.6	0.07968
0.4	0.6	0.83	0.5	0.0996
0.4	0.4	0.83	0.5	0.0664
0.2	0.6	0.83	0.55	0.05478
0.2	0.4	0.83	0.1	0.00664
		P(PSW=High | OW=medium)		**0.44654**

Calculating P(AD = High) from its table

AW	RTT	AD.High		AW	RTT	AD.High	
High	Yes	0.7		0.4	0.7	0.7	0.196
High	No	0.9		0.4	0.3	0.9	0.108
Medium	Yes	0.6	table becomes	0.4	0.7	0.6	0.168
Medium	No	0.7		0.4	0.3	0.7	0.084
Low	Yes	0.3		0.2	0.7	0.3	0.042
Low	No	0.5		0.2	0.3	0.5	0.03
						P(AD = High)	**0.628**

Hence P(AD = low) = 1 - P(AD = high) = 0.372

Hence we can now redraw the table for X=2, now with probabilities filled for each cell

AD	PSW	SL	AT	P(X2=YES)	
0.628	0.4456	0.32	1	0.65	0.0582060544
0.628	0.4456	0.51	1	0.65	0.0927658992
0.628	0.4456	0.17	1	0.5	0.023786128
0.372	0.4456	0.32	1	0.8	0.0424353792
0.372	0.4456	0.51	1	0.45	0.0380426544
0.372	0.4456	0.17	1	0.1	0.0028179744
					0.2580540896

Therefore the answer to our query is

0.2580540896