香 港 資 優 教 育 學 苑
The Hong Kong Academy for Gifted Education

Analysis of Telephone Survey Results on Student Satisfaction using Bayesian Network (BN) Approach with Latent Variables

Introduction

To further analyze the results of the telephone survey (conducted by Chinese University of Hong Kong in 2016) on student and alumni satisfactions and its relationship with their academic achievements and affective development, we represent the possible associations/dependence amongst question items and certain demographic information of members (e.g., study domain and study level) in form of a discrete Bayesian Network (BN), together with a number of latent variables. A Bayesian Network is a probabilistic graphical model (a type of statistical model) that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG). For example, a Bayesian Network could represent the probabilistic relationships between diseases and symptoms. Given symptoms, the network can be used to compute the probabilities of the presence of various diseases.

BN has been used successfully to model complex system in diverse fields, including ecology and environment. It has also been used in the past for customer satisfaction modelling (e.g., Salini and Kenett 2009; Hsu et al. 2009; Turkyilmaz et al. 2013). These studies have shown distinct advantages of BNs, including the ability to model complex interrelations between factors, perform scenario analysis, undertake sophisticated interrogations of the system, and include other sources of information in the model, such as results from published literature and expert judgement.

Formulation of Bayesian Network

In the analysis, the question items on the satisfaction of services and programmes respectively related to academic, affective (social interactions and personal growth), and overall aspects were focused. Besides, aspects related to academic outcomes (which were derived according to scholarships and competition prizes obtained) and affective outcomes (i.e., anything to be proud of recently) were concerned. Totally, there were 16 question items under consideration:

Academic dimension:

X1: Satisfaction in increasing knowledge in a specific field/ domain (Agreed/Neutral or Not Agreed)
X2: Satisfaction in increasing interest in a specific field/domain (Agreed/Neutral or Not Agreed)
X3: Satisfaction in improving study skills (Agreed/Neutral or Not Agreed)
X4: Satisfaction in improving academic results or development (Agreed/Neutral or Not Agreed)

Social interaction dimension:

X5: Satisfaction in strengthening social abilities (Agreed/Neutral or Not Agreed)

X6: Satisfaction in meeting friends of similar interests or abilities (Agreed/Neutral or Not Agreed)

X7: Satisfaction in meeting friends of various backgrounds (Agreed/Neutral of Not Agreed)

X8: Satisfaction in broadening the scope of social life (Agreed/Neutral or Not Agreed)

Personal growth dimension:

X9: Satisfaction in broadening the perspective (Agreed/Neutral or Not Agreed)

X10: Satisfaction in helping the formulation of future plan of development (Agreed/Neutral or Not Agreed)

X11: Satisfaction in strengthening your confidence (Agreed/Neutral or Not Agreed)

X12: Satisfaction in strengthening leadership abilities (Agreed/Neutral or Not Agreed)

Outcome:

X13: Academic performance level gauged by considering the performance in local and international competitions, and scholarship (Very High/High/Mid/Low)

X14: Some recent achievements to be proud of (Yes/No)

Overall Satisfaction:

X15: Overall satisfaction in the services and programmes related to personal growth and development (Agreed/ Neutral or Not Agreed)

X16: Overall satisfaction in the services and programmes delivered by the Academy (Agreed/Neutral or Not Agreed)

Demographic variable:

Dom: Domain (Sci & Math/ Humanities &Leadership/ Multi-domain)

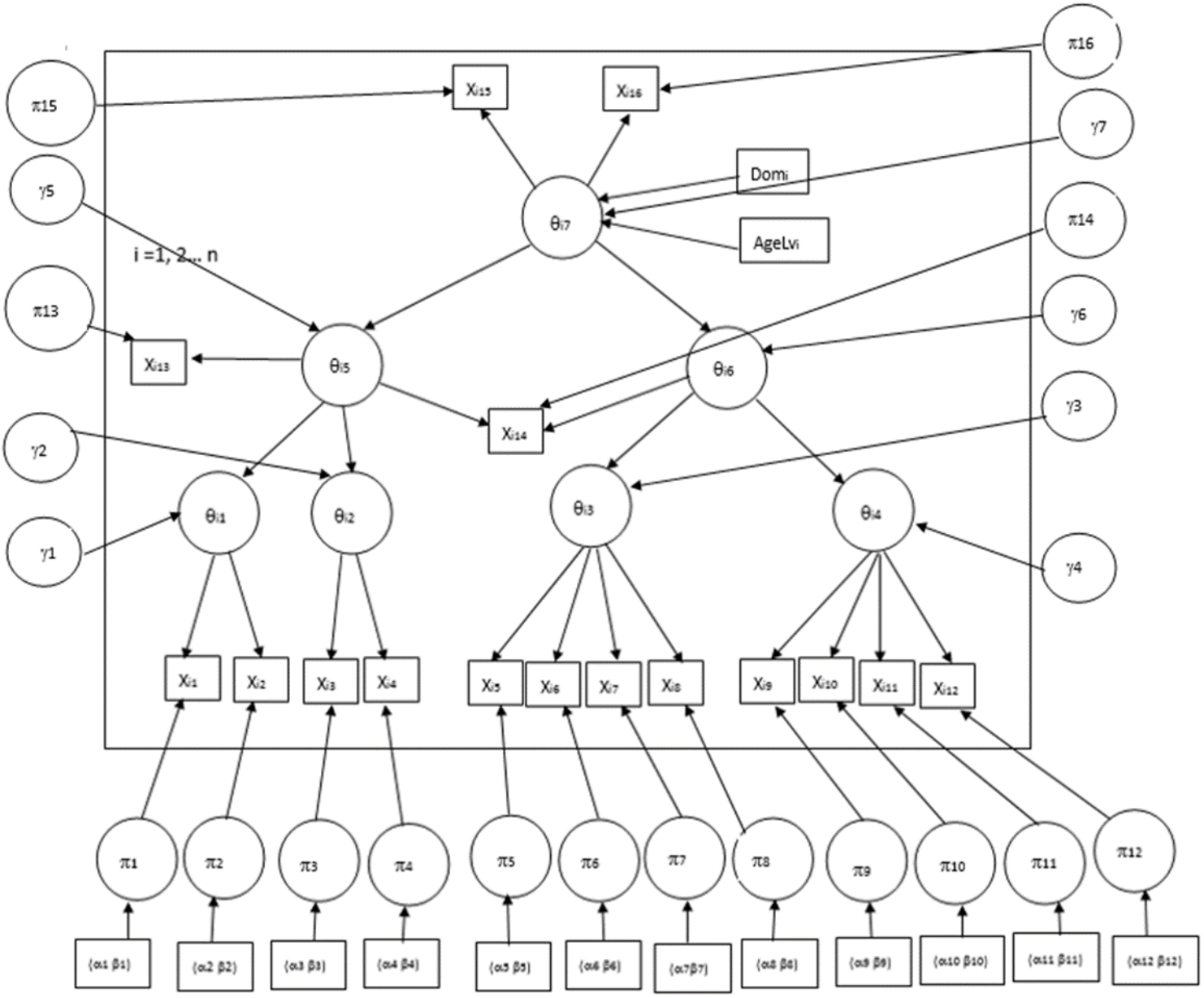
AgeLv: Age-Level (Junior Secondary/Senior Secondary/ Alumni)

Most of these questions were asked using a 5-point Likert Scale. Responses of 4 or 5 were classified as 'Agreed' with the statements concerned (encoded as 1); while responses of 3 or below were classified as 'Neutral or Not Agreed' with the statements concerned (encoded as 0).

Similar to the approach in Structural Equation Model (SEM), a number of observed responses are supposed to be due the impact of a latent variable. For provision of a concise representation, a hierarchical structure for latent variables is constructed. Together with some relevant demographic

information of students (i.e., study domain and study level), the resultant Bayesian network in the form of a DAG is formulated and shown below.

Figure 1: Directed Acyclic Graph (DAG) of Discrete Bayesian Network with Latent Variables for Survey Results (302 Secondary Student Members and 199 Alumni) conducted by Chinese University of Hong Kong in 2016



Notes: (i) The number, n runs from 1 to 501.
(ii) Square boxes represent observed/known variables (Xis); while circles represent latent/unknown variables (θis – Latent variables, πj and γj – unknown parameters of distributions) .
(iii) For the sake of clarity, only the known parameters of prior distributions for π1…π12 are displayed.

The interpretation of the latent variables is briefly described below.

θ_{i1}: The latent variable (representing satisfaction related to academic knowledge and interests) governs the responses of **X1** (Satisfaction related to academic knowledge) and **X2** (Satisfaction related to academic interests).

θ_{i2}: The latent variable (representing satisfaction related to study skills and academic results) governs the responses of **X3** (Satisfaction related to study skills) and **X4** (Satisfaction related to academic results).

θ_{i3}: The latent variable (representing satisfaction related to social interaction) governs the responses of **X5** (Satisfaction related to social abilities), **X6**(Satisfaction related to making friend with same interests), **X7**(Satisfaction related to making friends with various background), **X8**(Satisfaction related to broadening the scope of social life)**.**

θ_{i4}: The latent variable (representing satisfaction related to personal growth) governs the responses of **X9** (Satisfaction related to broadening perspectives), **X10**(Satisfaction related to future plan and development), **X11**(Satisfaction related to strengthening self-confidence), **X12**(Satisfaction related to leadership)**.**

θ_{i5}: The latent variable (representing overall satisfaction in academic aspects) governs the responses of **X13** (Academic performance level) and **X14** (Some recent achievements to be proud of), and the latent variables θ_{i1} and θ_{i2}.

θ_{i6}: The latent variable (representing overall satisfaction in affective aspects) governs the responses of **X14** (Some recent achievements to be proud of), and the latent variables θ_{i3} and θ_{i4}.

θ_{i7}: The latent variable (representing overall satisfaction) governs the responses of **X15** (Overall satisfaction related to personal growth and development) and **X16** (Overall satisfaction to services provided), and the latent variables θ_{i5} and θ_{i6}.

Corresponding to the DAG, the distributions of observed variables (X_i) and latent variables (θ_i) are specified accordingly.

Observed variables:

$(X_{i1}|\theta_{i1}=z,\pi_1) \sim$ Bernoulli(π_{1z}); for i=1, …n, z = 0,1

$(X_{i2}|\theta_{i1}=z,\pi_1) \sim$ Bernoulli(π_{2z}); for i=1, …n, z = 0,1

$(X_{i3}|\theta_{i2}=z,\pi_1) \sim$ Bernoulli(π_{3z}); for i=1, …n, z = 0,1

$(X_{i4}|\theta_{i2}=z,\pi_1) \sim$ Bernoulli(π_{4z}); for i=1, …n, z = 0,1

$(X_{i5}|\theta_{i3}=z,\pi_1) \sim$ Bernoulli(π_{5z}); for i=1, …n, z = 0,1

$(X_{i6}|\theta_{i3}=z,\pi_1) \sim$ Bernoulli(π_{6z}); for i=1, …n, z = 0,1

$(X_{i7}|\theta_{i3}=z,\pi_1) \sim$ Bernoulli(π_{7z}); for i=1, …n, z = 0,1

$(X_{i8}|\theta_{i3}=z,\pi_1) \sim$ Bernoulli(π_{8z}); for i=1, …n, z = 0,1

$(X_{i9}|\theta_{i4}=z,\pi_1) \sim$ Bernoulli(π_{9z}); for i=1, …n, z = 0,1

$(X_{i10}|\theta_{i4}=z,\pi_1) \sim$ Bernoulli(π_{10z}); for i=1, …n, z = 0,1

$(X_{i11}|\theta_{i4}=z,\pi_1) \sim$ Bernoulli(π_{11z}); for i=1, …n, z = 0,1

$(X_{i12}|\theta_{i4}=z,\pi_1) \sim$ Bernoulli(π_{12z}); for i=1, …n, z = 0,1

$(X_{i13}|\theta_{i5}=z,\pi_{13}) \sim$ Categorical($\pi_{13z,0}, \pi_{13z,1}, \pi_{13z,2}, \pi_{13z,3}$); for i=1, …n, z = 0,1

$(X_{i14}|(\theta_{i6}, \theta_{i5}) = (z1, z2), \pi_{14}) \sim$ Bernoulli($\pi_{14z1, z2}$); for i=1, …n, z1 = 0,1, z2=0,1

$(X_{i15}|\theta_{i7}=z,\pi_{15}) \sim$ Bernoulli(π_{15z}); for i=1, …n, z = 0,1

$(X_{i16}|\theta_{i7}=z,\pi_{16}) \sim$ Bernoulli(π_{16z}); for i=1, …n, z = 0,1

Latent variables:

$(\theta_{i1}|\theta_{i5} = z, \gamma 1) \sim$ Bernoulli(γ_{1z}); for i=1, …n, z=0,1

$(\theta_{i2}|\theta_{i5} = z, \gamma 2) \sim$ Bernoulli(γ_{2z}); for i=1, …n, z=0,1

$(\theta_{i3}|\theta_{i6} = z, \gamma 3) \sim$ Bernoulli(γ_{3z}); for i=1, …n, z=0,1

$(\theta_{i4}|\theta_{i6} = z, \gamma 4) \sim$ Bernoulli(γ_{4z}); for i=1, …n, z=0,1

$(\theta_{i5}|\theta_{i7} = z, \gamma 1) \sim$ Bernoulli(γ_{5z}); for i=1, …n, z=0,1

$(\theta_{i6}|\theta_{i7} = z, \gamma 2) \sim$ Bernoulli(γ_{6z}); for i=1, …n, z=0,1

$(\theta_{i7}|$Dom[i]=d, AgeLv[i]=a) $\sim$ Bernoulli($\gamma_{7d,a}$); for i=1, …n, Dom[i]=1,2,3 AgeLv[i]=1,2

Under the Bayesian framework, the unknown parameters shown in the above distributions of observed and latent variables are also regarded as variables. Their prior distributions are specified and presented below.

Uninformative Prior distributions of Unknown Parameters:

$\pi_{j0} \sim$ Beta (2, 8); for j =1,2 … 12,15,16

$\pi_{j1} \sim$ Beta(8, 2); for j =1,2 … 12,15,16

$\pi_{140,0} \sim$Beta(2, 8)

$\pi_{140,1} \sim$Beta(5, 5)

$\pi_{141,0} \sim$Beta(5, 5)

$\pi_{141,1} \sim$Beta(8, 2)

$\pi_{130} \sim$ Dirchlet(4,3,2,1)

$\pi_{131} \sim$ Dirchlet(1,2,3,4)

$\gamma_{j0} \sim$ Beta (2, 8); for j =1,2 … 6

$\gamma_{j1} \sim$ Beta (8, 2); for j =1,2 … 6

$\gamma 7_{i,j} \sim$ Beta (5, 5); for i=1,2,3 ,j =1,2

It should be noted that the actual real numbers used in the prior distributions are relatively small, as compared with the data samples (n = 502). Thus, the prior distributions could be regarded as uninformative, and the estimation results would be data driven.

Estimation Method and Analysis of Results

Estimation of Posterior Distributions of Latent Variables and Parameters using MCMC

Under Bayesian framework, the difficult part is to estimate the posterior distribution of the unknowns (latent variables and parameters of distributions) given the data and the prior distributions. Recently, the freeware WINBUGS has became widely popular as it can estimate the posterior distributions of the unknowns in a variety of models using the simulation techniques, Markov Chain Monte Carlo (MCMC). It only requires to specify the model code in which the model likelihood and the prior distribution are defined. The estimates of latent variables and parameters concerned, which could be derived using the simulation approach, are discussed below.

Bayesian Analysis of Latent Variables and Unknown Parameters using Posterior Distributions

First, we examine the posterior distributions of the averages (over all respondents) of latent variables, of which the corresponding summary statistics are presented below.

Table 1: Summary statistics for the estimations of the averages of latent variables

Average	mean	sd	MC error	2.5%	median	97.5%	start	sample
average θi1	0.6713	0.03094	0.002364	0.6068	0.6727	0.7305	1001	4000
average θi2	0.5208	0.04331	0.003288	0.4371	0.5190	0.6088	1001	4000
average θi3	0.5179	0.02286	0.001494	0.4711	0.5190	0.5609	1001	4000
average θi4	0.5687	0.02269	0.001782	0.5210	0.5709	0.6088	1001	4000
average θi5	0.6285	0.02271	0.002181	0.5808	0.6287	0.6727	1001	4000
average θi6	0.5866	0.02105	0.00183	0.5429	0.5868	0.6248	1001	4000
average θi7	0.6223	0.01961	0.001962	0.5828	0.6228	0.6607	1001	4000

From **Table 1**, the followings can be observed:

- The average of θi7, reflecting the overall satisfaction, amounted to 0.62. The latent variables affected the responses of X15 (Overall satisfaction related to personal growth and development, and the proportion of respondents answering 'Agreed' = 0.62 and X16 (Overall satisfaction related to the programmes and services provided by the Academy, and the proportion of 'Agreed' =0.75). Besides, it is found that the overall satisfaction level (θi7) decreased, as the study level (junior secondary/ senior secondary/ alumni) of student members increased (averages of θi7 for junior

secondary, senior secondary and alumni are respectively 0.76, 0.57, 0.50). This remained true for each individual domain.

- Overall speaking, the programmes and services related to academics were rated higher than those related to affective development, as the average of θ_{i5} (0.63) is greater than the average of θ_{i6} (0.59).

- Within the academic aspect, the respondents were more satisfied with the impacts of the programmes and services to their academic knowledge and interests (average of θ_{i1} = 0.67), as compared with the impacts to their study skills and academic results (average of θ_{i2} = 0.52).

  This was in line with the proportions of 'Agreed' respectively for related question items (proportions of 'Agreed' for X1, X2, X3 and X4 are respectively 0.75, 0.72, 0.38, and 0.40).

- With the aspect of affective development, the respondents were slightly more satisfied with the impacts of the programmes and services to their personal growth, as compared with the impacts to their social interactions; since the average of θ_{i4} = 0.57 was slightly greater than the average of θ_{i3} = 0.52.

  This was quite in line with the proportions of 'Agreed' respectively for related question items (proportions of 'Agreed' for X5, X6, X7 and X8 are respectively 0.44, 0.55, 0.52, and 0.46, and proportions of 'Agreed' for X9, X10, X11 and X12 are respectively 0.79, 0.43, 0.49, and 0.37).

Besides, we examine the estimates of some parameters, namely π_{13} and π_{14} (see **Table 2a** and **Table 2b**).

Table 2a: Summary statistics for the estimations of the parameter π_{13}

parameter	mean	sd	MC error	2.5%	median	97.5%	start	sample
$\pi_{13_{0,0}}$	0.2735	0.03426	9.549E-4	0.2081	0.2734	0.3421	1001	4000
$\pi_{13_{0,1}}$	0.3353	0.03713	0.001216	0.2606	0.3354	0.4093	1001	4000
$\pi_{13_{0,2}}$	0.2901	0.03541	0.001072	0.2238	0.2885	0.3619	1001	4000
$\pi_{13_{0,3}}$	0.1012	0.02466	0.001092	0.05581	0.09996	0.1515	1001	4000
$\pi_{13_{1,0}}$	0.2408	0.02484	6.278E-4	0.1933	0.2401	0.2921	1001	4000
$\pi_{13_{1,1}}$	0.2964	0.02758	9.131E-4	0.2433	0.2964	0.3508	1001	4000
$\pi_{13_{1,2}}$	0.2493	0.02507	6.44E-4	0.2026	0.2485	0.3009	1001	4000
$\pi_{13_{1,3}}$	0.2135	0.02361	6.287E-4	0.1672	0.213	0.2602	1001	4000

Table 2b: Summary statistics for the estimations of the parameter π_{14}

parameter	mean	sd	MC error	2.5%	median	97.5%	start	sample
$\pi 14_{0,0}$	0.4863	0.04093	0.001434	0.4055	0.4868	0.5656	1001	4000
$\pi 14_{0,1}$	0.5940	0.1369	0.004725	0.3163	0.6004	0.8432	1001	4000
$\pi 14_{1,0}$	0.4287	0.1143	0.005413	0.2063	0.4274	0.6497	1001	4000
$\pi 14_{1,1}$	0.7523	0.02774	7.559E-4	0.6954	0.752	0.8044	1001	4000

$\pi 13_0$ is the vector of estimated probabilities that a student/alumnus attained various levels of academic performance (Low/Middle/High/Very High) given that he/she was not overall satisfied on the academic aspect (i.e. $\theta i5=0$); vice versa for $\pi 13_1$. From **Table 2a**, it is obvious that those students/ alumni who were overall satisfied on the academic aspects had much higher chance of obtaining "Very High" academic performance ($\pi 13_{1,3}$ = 0.2135), as compared with those who were not overall satisfied on academic aspect ($\pi 13_{0,3}$= 0.1012). Such a phenomenon hints that the learning opportunities provided by the Academy may help the students' academic performance in their studies.

$\pi 14$ is the estimated probabilities that a student/alumnus would respond to 'Yes' when he/she was asked whether he/she had any recent achievements that he/she was proud of. $\pi 14_{0,0}$ represents the estimated probability when he/she was not satisfied on both academic on affective aspects; while $\pi 14_{0,1}$ represents the estimated probability when he/she was not satisfied academic aspect; but was satisfied with affective aspect. The interpretations for $\pi 14_{1,0}$ and $\pi 14_{1,1}$ are similar. From **Table 2b**, it may be interesting to note that the probability of responding 'Yes' was higher for those who were solely satisfied with the affective aspect ($\pi 14_{0,1}$ = 0.59) than those who were solely satisfied with academic aspect ($\pi 14_{1,0}$ = 0.43). It may imply that affective development may have greater impact to students of being able to feel proud of themselves than academic achievements.

What-if Analysis using Probabilistic Logic Programming (Problog)

Last, we would conduct a kind of what-if analysis to examine the expected outcomes under different scenarios. The following three scenarios will be explored:

Scenario 1: With the improvement on programmes and services related to personal growth such that all responses to related question items (X9, X10, X11 and X12) are all positive, what will be the expected outcomes of other interested variables?

Scenario 2: With the improvement on programmes and services related to social interactions such that all responses to related question items (X5, X6, X7 and X8) are all positive, what will be the expected outcomes of other interested variables?

Scenario 3: With the improvement on programmes and services related to academic

knowledge and interests such that all responses to related question items (X1, and X2) are all positive, what will be the expected outcomes of other interested variables?

With regard to expected outcomes, we would examine the following variables under various scenarios.

- Resultant probabilities of being able to feel proud of some recent achievements (pr_X14)
- Resultant probabilities of feeling overall satisfied on the impacts of programmes and services to personal growth and development (pr_X15)
- Resultant probabilities of feeling overall satisfied on the programmes and services provided (pr_X16)

To achieve this kind of what-if analysis, there a number of tools available. In this study, we adopt the framework of probabilistic logic programming; in particular, we use the programming language, called ProbLog. Probabilistic logic programs are logic programs in which some of the facts are annotated with probabilities. ProbLog is a tool that allows uses to intuitively build programs that could encode complex interactions between a large set of heterogenous components and express the inherent uncertainties that are present in real-life situations. It is a suite of efficient algorithms for various inference tasks (e.g., computing the marginals given evidence). It is based on a conversion of the program and the queries and evidence to a weighted Boolean formula. This allows users to reduce the inference tasks to well-studied tasks, which can be solved using state-of-the-art methods known from the graphical model and knowledge compilation literature. For further details, please refer to Daan Fierens et al. 2015.

The BN shown in **Figure 1** above could be encoded in Problog as shown in **Annex 1**. By running appropriate queries with different provisions of evidences corresponding to various scenarios, the resultant probabilities of interested variables could be obtained and shown in **Table 3** below.

Table 3: What-if analysis: Expected outcomes under different scenarios

Scenario\Outcome	pr_X14	pr_X15	pr_X16
Baseline	0.63	0.60	0.73
Scenario 1	0.73	0.86	0.89
Scenario 2	0.70	0.78	0.84
Scenario 3	0.68	0.76	0.83

Relatively speaking, from **Table 3** it can be observed that improvement of programmes and

services related to personal growth could bring out the most prominent impacts to student members.

Posterior Predictive Model Checking (PPMC)

In Bayesian framework, posterior predictive p-value based on a goodness-of-fit testing statistic is a common approach to check a model, which is derived as follows.

We consider the discrepancy measure of person fit. The squared Pearson residual for examinee i and observed response j is adopted and is defined as follows.

$$V_{ij}(x_{ij}, \boldsymbol{\theta}_i, \boldsymbol{\pi}_j) = \frac{(x_{ij} - P_{ij})^2}{P_{ij}(1 - P_{ij})}$$

$$P_{ij} = E(x_{ij} \,|\theta_i, \pi_j) \qquad \text{where } i = 1\ldots501, j=1\ldots12$$

Here the first 12 question items (X1,…,X12) are concerned. A person fit discrepancy measure could then be given by the root mean square error taken with respect to the values; i.e., for examinee i, we have

$$PF_i(\mathbf{x}_i, \boldsymbol{\theta}, \boldsymbol{\pi}_j) = \left(\frac{1}{J} \sum_{j=1}^{J} V_{ij}(x_{ij}, \boldsymbol{\theta}_i, \boldsymbol{\pi}_j) \right)^{\frac{1}{2}}$$

For each individual, i, we calculate in each simulation run $PF_i(x_i, \theta, \pi_j)$ using the actual observations x_i and the simulated data x_i^{rep}, $PF_i(x_i^{rep}, \theta, \pi_j)$. We compare the values of these two statistics and estimate the following by repeating the procedures a large number of times.

$$p_{Bi} = Pr(\, PF_i(x_i, \theta, \pi_j) \geq PF_i(x_i^{rep}, \theta, \pi_j) \,|\text{observed data}\,)$$

When p_{Bi} is not too far away from 0.5, (e.g., p_{Bi} >0.05 or p_{Bi} <0.95), we could conclude that the model fits the observations of person i. We have compiled p_{Bi} for all student members and alumni in the study. Most of their p_{Bi} fell in the acceptable range (only 7.4% of the respondents with their p_{Bi} being less than 0.05 or greater than 0.95). For instances, the students S1 and S2 had p_{Bi} being respectively equal to 0.42 and 0.55. The comparison of these two students' actual responses with the corresponding simulated responses were shown in **Table 4** below.

Table 4: Summary statistics of 12 question items for model checking

Question	S1 (p-value=0.42)		S2 (p-value=0.55)		Proportion Of "Agreed"	
	Observed	Average over 4000 runs	Observed	Average over 4000 runs	Observed	Est. Values based on Est. Parameters (π_j & γ_j)
1	1	0.9445	1	0.9483	0.7545	0.7352
2	1	0.8638	1	0.8710	0.7166	0.7002
3	1	0.5917	1	0.5998	0.3772	0.3741
4	1	0.6697	1	0.6603	0.4012	0.4001
5	1	0.6717	1	0.2400	0.4411	0.4391
6	1	0.8515	0	0.2645	0.5469	0.5403
7	0	0.8177	0	0.2352	0.5210	0.5155
8	1	0.8273	0	0.1115	0.4631	0.4577
9	1	0.9433	1	0.9060	0.7964	0.7846
10	1	0.6432	0	0.5885	0.4271	0.4254
11	1	0.7742	1	0.7037	0.4910	0.4839
12	1	0.5972	0	0.5305	0.3673	0.3647

From Table 4, it can be noted that the observed responses of S1 and S2 are in line with the corresponding averages derived from 4000 simulation runs. In addition, we display the proportion of 'Agreed' for these 12 questions and compare them with the corresponding estimated values based on the estimated parameters (π_j and γ_j). A close match between these two sets of values is discerned.

Summary of Findings

In the study, we analyze the survey results related to student satisfaction, which was conducted by Chinese University of Hong Kong in 2016. The main findings are summarized below:

- Overall speaking, the programmes and services related to academics were rated higher than those related to affective development.

- Within the academic aspect, the respondents were more satisfied with the impacts of the programmes and services to their academic knowledge and interests, as compared with the impacts to their study skills and academic results.

- Those students/ alumni who were overall satisfied on the academic aspects had much higher chance of obtaining "Very High" academic performance. It may hint that the learning opportunities provided by the Academy may help the students' academic performance in their studies.

- The study results may imply that affective development may have greater impact to students of being able to feel proud of themselves than solely academic achievements.

- In what-if analysis, that improvement of programmes and services related to personal growth could bring out the most prominent impacts to student members, as compared with the ones respectively related to social interaction and academic enhancement.

References

Daan Fierens, Guy Van den Broeck, Joris Renkens, Dimitar Shterionov, Bernd Gutmann, Ingo Thon, Gerda Janssens, and Luc De Raedt. Inference and learning in probabilistic logic programs using weighted Boolean formulas. *Theory and Practice of Logic Programming*, 2015.

Hsu C-I, Shih M-L, Huang B-W, Lin B-Y, Bin C-N. Predicting tourism loyalty using an integrated Bayesian network mechanism. Exp. Syst Appl. 2009; 36(9):11760-3. https://doi.org/10.1016/j.eswa.2009.04.010

Salini S, Kenett RS. Bayesian networks of customer satisfaction survey data. *J. Appl. Stat.*, 2009; 36(11): 1177-89.

Turkylimaz A, Oztekin A, Zaim S, Demirek OF. Universal structure modeling approach to customer satisfaction index. Ind Manag Data Syst. 2013; 113(7): 932-49.

Annex 1

```
% The model in Probabilistic Logic Program (ProbLog)
0.18::sci_junsec;0.19::sci_sensec;0.227::sci_alumni;0.11::hum_junsec;0.077::hum_sensec;
0.16::hum_alumni;0.020::multi_junsec; 0.022::multi_sensec;0.014::multi_alumni.

0.76::theta_over_sat :- sci_junsec.
0.55::theta_over_sat :- sci_sensec.
0.50::theta_over_sat :- sci_alumni.
0.78::theta_over_sat :- hum_junsec.
0.62::theta_over_sat :- hum_sensec.
0.49::theta_over_sat :- hum_alumni.
0.69::theta_over_sat :- multi_junsec.
0.60::theta_over_sat :- multi_sensec.
0.50::theta_over_sat :- multi_alumni.

0.90::q15_over_pgd :- theta_over_sat.
0.16::q15_over_pgd :- \+theta_over_sat.

0.92::q16_over_ser :- theta_over_sat.
0.44::q16_over_ser :- \+theta_over_sat.

0.97::theta_acd_over :- theta_over_sat.
0.061::theta_acd_over :- \+theta_over_sat.

0.92::theta_aff_over :- theta_over_sat.
0.041::theta_aff_over :- \+theta_over_sat.

0.75::q14_proud_of :- theta_acd_over, theta_aff_over.
0.59::q14_proud_of :- \+theta_acd_over, theta_aff_over.
0.43::q14_proud_of :- theta_acd_over, \+theta_aff_over.
0.49::q14_proud_of :- \+theta_acd_over, \+theta_aff_over.

0.24::acp_l;0.30::acp_m;0.25::acp_h;0.21::acp_vh :- theta_acd_over.
0.27::acp_l;0.34::acp_m;0.29::acp_h;0.10::acp_vh :- \+theta_acd_over.

0.96::theta_acd_knIn :- theta_acd_over.
0.18::theta_acd_knIn :- \+theta_acd_over.
0.81::theta_acd_stdex :- theta_acd_over.
0.044::theta_acd_stdex :- \+theta_acd_over.

0.75::theta_aff_soc :- theta_aff_over.
0.19::theta_aff_soc :- \+theta_aff_over.
0.94::theta_aff_pg :- theta_aff_over.
0.042::theta_aff_pg :- \+theta_aff_over.

0.95::q1_acd_kn :- theta_acd_knIn.
0.33::q1_acd_kn :- \+theta_acd_knIn.
0.87::q2_acd_In :- theta_acd_knIn.
0.38::q2_acd_In :- \+theta_acd_knIn.
0.60::q3_acd_std :- theta_acd_stdex.
0.14::q3_acd_std :- \+theta_acd_stdex.
0.68::q4_acd_exam :- theta_acd_stdex.
```

```
0.11::q4_acd_exam :- \+theta_acd_stdex.

0.67::q5_soc_Int :- theta_aff_soc.
0.20::q5_soc_Int :- \+theta_aff_soc.
0.84::q6_soc_fin :- theta_aff_soc.
0.23::q6_soc_fin :- \+theta_aff_soc.
0.83::q7_soc_fbk :- theta_aff_soc.
0.19::q7_soc_fbk :- \+theta_aff_soc.
0.84::q8_soc_cir :- theta_aff_soc.
0.062::q8_soc_cir :- \+theta_aff_soc.

0.95::q9_pg_pre :- theta_aff_pg.
0.58::q9_pg_pre :- \+theta_aff_pg.
0.64::q10_pg_plan :- theta_aff_pg.
0.16::q10_pg_plan :- \+theta_aff_pg.
0.77::q11_pg_conf :- theta_aff_pg.
0.13::q11_pg_conf :- \+theta_aff_pg.
0.59::q12_pg_lead :- theta_aff_pg.
0.086::q12_pg_lead :- \+theta_aff_pg.

%scen1 :- q9_pg_pre, q10_pg_plan, q11_pg_conf, q12_pg_lead.
%evidence(scen1, true).
%scen2 :- q5_soc_Int, q6_soc_fin, q7_soc_fbk, q8_soc_cir.
%evidence(scen2, true).
%scen3 :- q1_acd_kn, q2_acd_In.
%evidence(scen3, true).

query(q14_proud_of).
query(q15_over_pgd).
query(q16_over_ser).
```