Use of Bayesian Network and Markov Network in drawing inference and its application in Smartphone keystrokes minimization

A.N. Patil ankitnar@buffalo.edu

Abstract

Bayesian Networks and Markov Networks are widely helpful in making inference and understanding the flow of influence. We discuss their interesting application to the children handwriting dataset and external datasets like Male Fertility [1] and Car Acceptability [2] datasets from UCI Machine Learning repository. We can easily identify most influencing factors called primary factors and less influencing factors called secondary factors that influence the final outcome. Another interesting application of Bayesian Network in predicting the next word in a Smartphone based chat application to minimize the number of keystrokes is also discussed. This approach is language independent and personalized. Private chat messages from WhatsApp messenger [7] were used for this experiment. Also the use of Deep Belief network in classification of handwritten numbers is discussed in the last section. Various inference algorithms are applied in on the children handwriting datasets and inferences are drawn.

1. Data Description and Data Cleaning

2011-12 First Grade printing dataset and 2012-13 Second Grade printing dataset were used for analysis and comparison. 2011-12 First Grade printing dataset is referred to as Dataset1 and 2012-13 Second Grade printing dataset is referred to as Dataset2 in the following report. X1 – X12 represents features in the Datasets.

Missing features in the data set can be estimated, given that number of missing features is low as compared to total number of features in the sample. In order to estimate the missing features, the features that are present in the sample are used. We find clean samples that closely match the features in the dirty sample and use the features in the clean sample to estimate the missing features.

For example consider following example of dirty data sample and another similar data sample. In data sample R1 the missing feature in X12. We find a data sample which is similar to R1 from the data set which is R2. Now the missing feature X12 can be estimated to be same as that of R2 since most of the other features are similar to R1.

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
R1	0	0	0	1	1	1	3	0	99	2	2	**99**
R2	1	0	0	1	0	1	3	0	99	1	2	**1**

Table 1: Data cleaning using similar data sample

2. Inference Algorithm

2.1 Mean

Mean gives the concentration of features in the samples. Mean for each feature set was determined for both the datasets and compared.

Inference: Feature X1 and X2 has means closer to each other in both the datasets. Their distributions may be identical. If they have similar distributions, we can conclude that these features are less likely to change with as the child progresses.

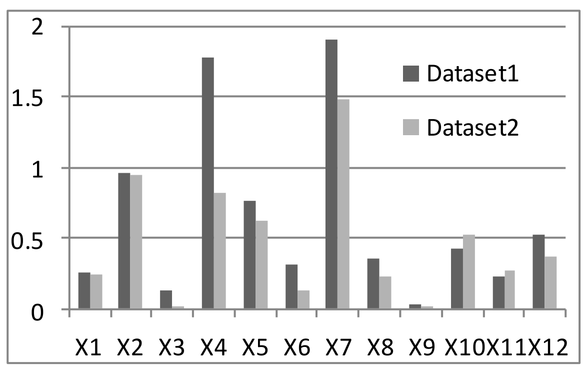


Figure 1: Mean

2.2 Entropy

Entropy will give the measure of the variations of features in the individual dataset. Comparison between entropy of features in two datasets will give the comparison of variations in the two datasets.

Inference: Features X2, X4 and X7 has similar entropy for both the dataset indicating that variations in these features remained constant over the year. The entropy of features X10, X11 and X12 has decreased for the Dataset2, indicating that these features are more likely to be same as the child progresses.

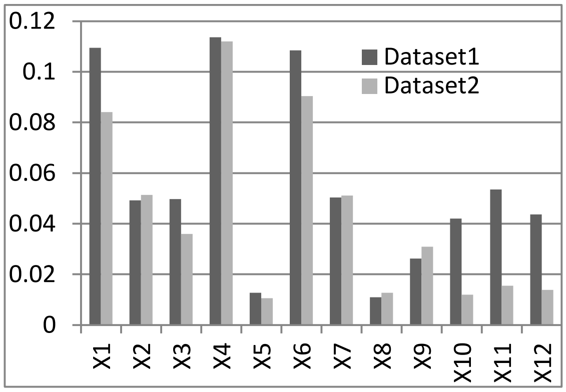


Figure 2: Entropy

2.3 Kullback-Leibler divergence

KL Divergence or relative entropy gives the similarity measure for two distributions. Higher the KL divergence greater the dissimilarity between two distributions.

Inference: Features X4, X10, X11 and X12 have larger values hence we can conclude that there is a wide difference in probability distributions between two datasets. For rest of the features, the distributions don't vary much.

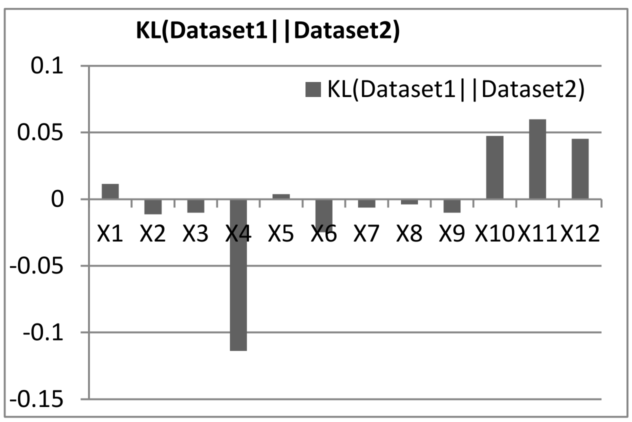
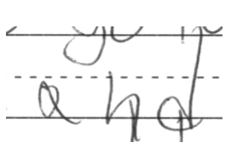


Figure 3: Kulback-Leibler divergence

2.4 Unusual writing

Most unusual writing can b determined by calculating the marginal probabilities of the characteristics. The sample set that gives the least probability of the features occurring is the most unusual writing in the provided dataset. Using the marginal probabilities the most unusual sample is {3,4,1,5,1,1,3,0,1,0,0,2}



3. Bayesian Network:

Pearson's chi-square [3][4] test was used to find the correlation between the nodes and log-loss was used to find the directionality of edges and to limit network size.

Measure of deviance (Pearson's chi-square test)

Chi-squared value was determined for all possible combinations of nodes. These values were ordered in descending manner to give the most likely candidate of pair of nodes in the BN construction.

$$\chi^2(\mathcal{D}) = \sum_{i,j} \frac{(O[x_i,y_j]-E[x_i,y_j])^2}{E[x_i,y_j]}.$$

Goodness of network (log-loss)

Log-loss was determined by taking the conditional probability of the candidate child node given the candidate parent node. This condition was used to determine the directionality of influence between nodes.

Limitations on parent node

In order to limit the number of edges in the BN and to reduce the complexity, a bound was set on number of in-degrees on a node. A node was not assigned parents greater than 2 irrespective of its log-loss value.

$$s(\mathcal{D}|G) = -\sum_{i=1}^{N}\sum_{j=1}^{n} \log P(x_{ij}|\mathrm{pa}_{ij}),$$

Efficiency (speed)

Chi-square takes $O(n^2)$ to find most capable candidate pair among all the nodes. The construction of BN can be further optimized by considering only those nodes that increases the log-loss. Efficiency is also increased by limiting the number of parents per node.

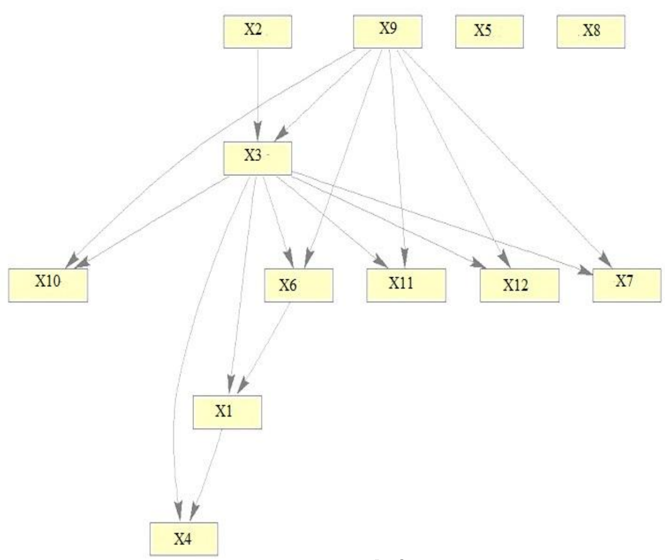


Figure 4: Bayesian Network for Dataset 1

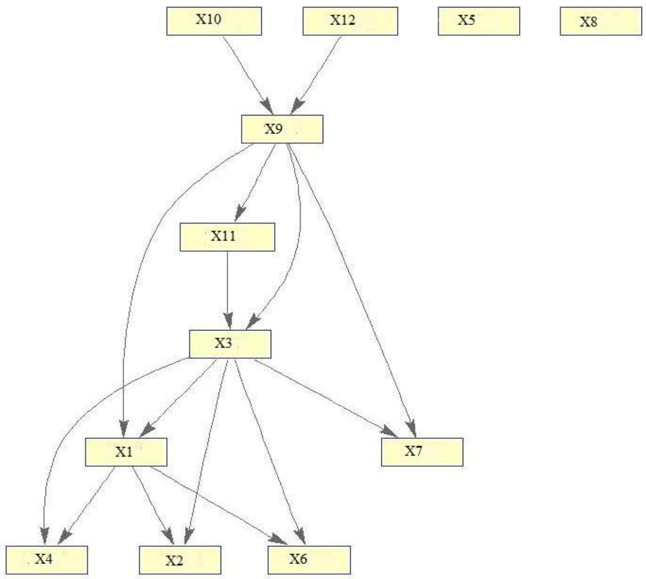


Figure 5: Bayesian Network for Dataset 2

4. Bayesian Network for Smart Phone keystrokes minimization

Need: Next word prediction on smart phones is crucial because it take efforts to type the whole word on smart phones due to small keyboard area. Since it is difficult to type on a smart phone, this approach tends to minimize the total number of key strokes. Also traditional dictionary bases approach don't work because of the short forms used. These short forms are highly personalized hence they vary from person to person. Also all the words written in English may not belong to the same language. Following diagram illustrates a small network with the nodes representing words and edges representing the probability of the next word. For instance if the user types W1, the next word would show W2, W3 and W4 as the possible next word depending on the probability. The user makes selection of the next word or types in the required word. Since the suggestions include most frequently following words, the number of key strokes in completing the sentence would be significantly reduced. By updating the learning probabilities this approach improves with time.

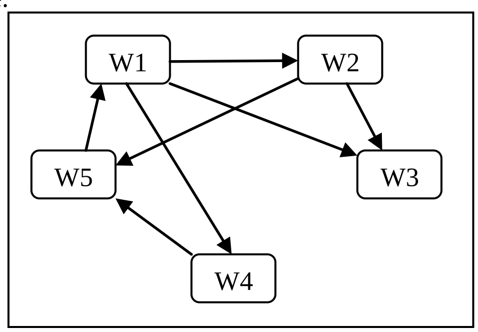


Figure 6: Sample word precedence graph

Dataset: Message history from WhatsApp chat application on smart phones was collected to analyze the word sequence and use it for predicting the next word. This chat uses English alphabets to represent Hindi words. Most of the chat is in Hindi but it also includes English and Marathi words.

Since the size of the Bayesian Network is huge in this case, only the next word probability for few words is shown.

INITIAL WORD	FIRST SUGGESTION	PROB.	2ND SUGGESTION	PROB.
KYA	BAAT	0.136	HE	0.306
			HAI	0.244
			KAR	0.163
	HUA	0.110	HE	0.304
			THA	0.173
			HAI	0.173
	KAR	0.063	RAHA	0.349
			RAHE	0.157
			RAHI	0.075

Table 2: Next word suggestion probabilities for Hindi

INITIAL WORD	1ST SUGG.	PROB.	2ND SUGG.	PROB.
WE	DON'T	0.146	KNOW	0.227
			WANT	0.136
			MAKE	0.136
	ARE	0.146	THE	0.138
			YOU	0.106
			A	0.095
	WILL	0.122	BE	0.381
			GET	0.090
			MISS	0.072

Table 2: Next word suggestion probabilities for English

Phrases like "*kya baat he*", "*kya hua tha*", "*kya kar raha*" can be easily be types with minimum key strokes and can be used for suggestion of the following words. English phrases like *"We don't know", "we will get", "we will miss"* can also be predicted using the same model.

5. Markov Network

Markov Network is constructed from sample set of observations. This is achieved by minimizing the Kulback-Leibler [5] divergence between the network under construction. This is carried out until a preset threshold is reached.

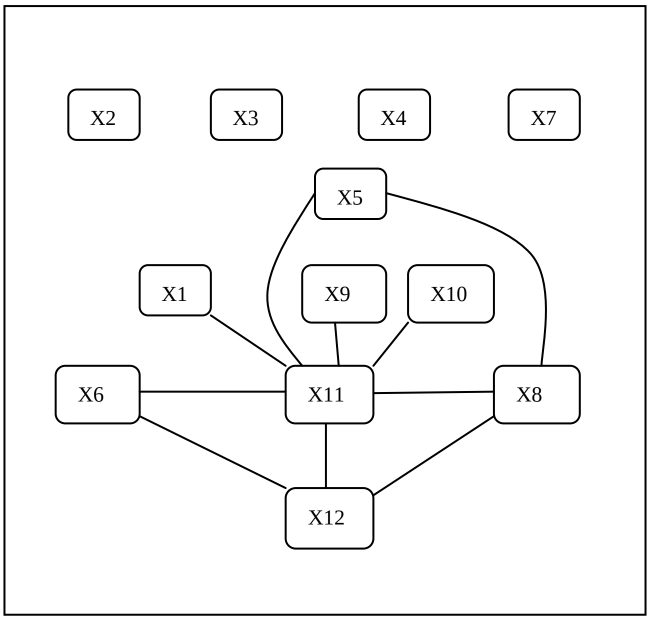


Figure 7: Markov network for Dataset 1

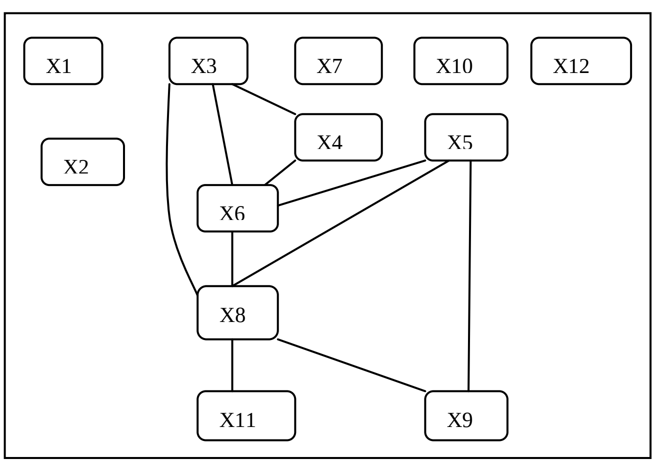


Figure 8: Markov network for Dataset 2

6. Markov Network on external datasets

6.1 Male fertility diagnosis

Dataset description: 100 data samples related to male fertility are used for this inference.

SEASON IN WHICH ANALYSIS WAS PERFORMED
AGE
CHILD DISEASES
ACCIDENT
SURGICAL INTERVENTION
HIGH FEVER IN LAST YEAR
FREQ. OF ALCOHOL CONSUMPTION
SMOKING HABIT
HOURS SPENT SITTING PER DAY
DIAGNOSIS NORMAL OR ALTERED

Table 3: Male fertility diagnosis data description

Inference:
<u>Least significant factors:</u> Few factors were found least significant in determining the final diagnosis. Frequency of alcohol consumption, smoking habit, season of the test and number of hours spent sitting were found least relevant in determination of diagnosis.
<u>Primary factors:</u> Surgical intervention, child disorder and Accident are connected by strong nodes with each other and the Diagnosis node. Hence we can conclude that these factors are more likely to determine the diagnosis result.
<u>Secondary factors:</u> High fever during last year and current age did not directly influence the diagnosis. But these factors have a high connectivity with all the primary factors. Hence these factors are also influential in determining the Diagnosis results.

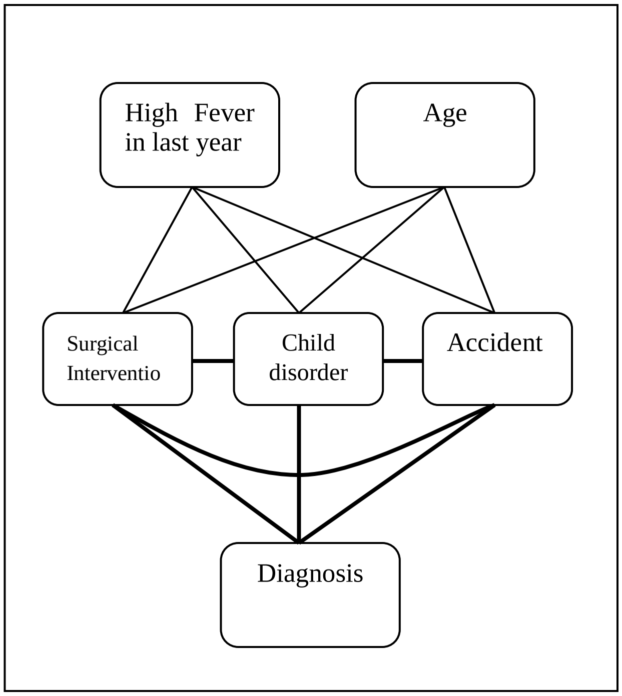


Figure 9: Markov Network for Male fertility diagnosis

6.2 Car acceptability

Dataset description: 1728 data samples related to various attributes related to car purchasing and acceptability are used for this inference.

CAR ACCEPTABILITY
BUYING PRICE
MAINTENANCE CHARGES
NO. OF DOORS
CAPACITY (# OF PERSONS)
SIZE OF LUGGAGE BOOT
SAFETY

Table 4: Car acceptability data description

Inference:
Primary factors: Safety, size of luggage boot and passenger capacity are connected by strong nodes with each other and the Acceptability node. Hence we can conclude that these factors are more likely to determine the overall acceptability of the car.

Secondary factors: Cost of maintenance, Number of doors and buying price did not directly influence the diagnosis. But these factors have a high connectivity with all the primary factors. Hence these factors are also influential in determining the Acceptability of car.

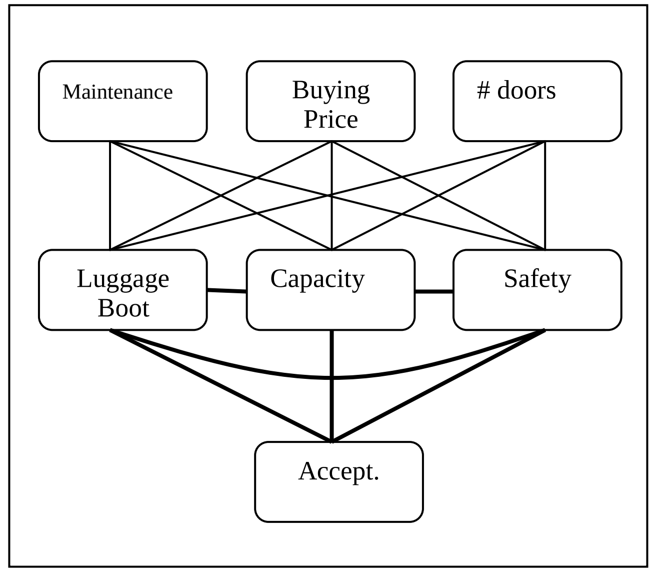


Figure 10: Markov Network for Car acceptability

7. Deep Belief Networks

Handwritten images of digits were classified using Deep Belief Networks. A Deep Learning Toolbox for Matlab [6] was used to accomplish this. A 100 hidden unit Restricted Boltzmann Machine was used to learn the weights.

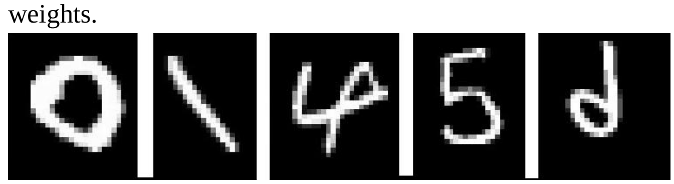


Figure 11: Sample images used for classification.

A high accuracy of 1.5 % was achieved for the classification purpose. The weights learnt can easily be related to the shapes of the digits.

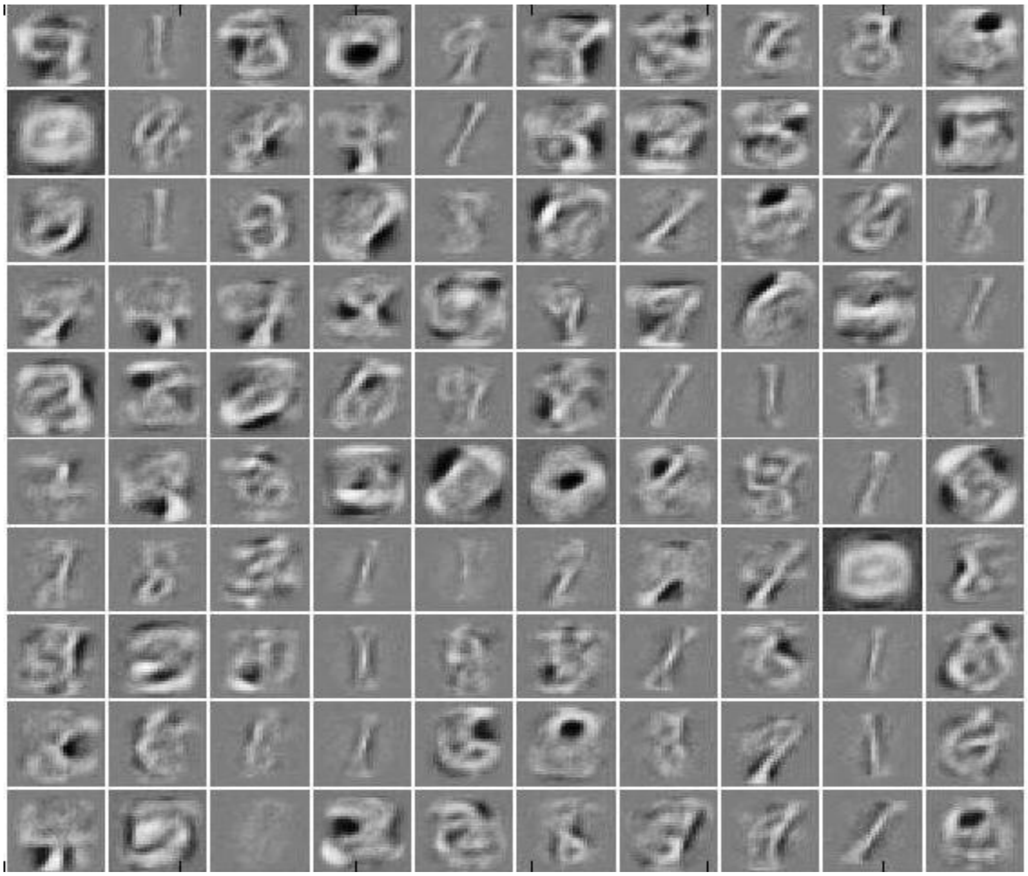


Figure 11: Weights learnt using RBM

8. Conclusion

Bayesian Networks and Markov Networks can be used to model the flow of influence and determine association of factors. We can easily identify most influential factors and their association with each other. On the other hand it also helps us to identify lest influential factors. This knowledge can be used for decision making and predictions. Causal analysis over a complex network can be achieved by tuning the connectivity of the network. Deep Learning is on the frontier of research in Artificial Intelligence. Excellent results can be achieved by using Deep Belief Networks.

9. References

[1] "Fertility Dataset",

http://archive.ics.uci.edu/ml/datasets/Fertility

[2] "Car Evaluation Dataset", https://archive.ics.uci.edu/ml/datasets/Car+Evaluation

[3] Mukta Puri, Sargur N. Srihari and Yi Tang, Bayesian Network Structure Learning and Inference Methods for Handwriting", 2013

[4] Yi Tang and Sargur N. Srihari, Efficient and Accurate Learning of Bayesian Networks using Chi-Squared Independence Tests, ICPR 2012

[5] S.K.M Wong and Y. Xiang, Construction of a Markov Network from data for probabilistic inference

[6] "Deep Learning Toolbox", http://www.mathworks.com/matlabcentral/fileexchange/38310-deep-learning-toolbox.

[7] "WhatsApp Messanger", http://www.whatsapp.com/