

# Probabilistic Analysis of Children's handwriting using Bayesian and Markov Networks

Aravind Kumar Ramesh (UB Person #: 5013- 2610)

## 1. ABSTRACT

Probabilistic graphical models aim to express the conditional dependencies in a structural hierarchical fashion among random variables. A graph based approach is used to model a multidimensional data; the factorized representation in itself takes care of the dependencies that exist within the multidimensional space of the data set. Bayesian networks encode factorized form of the joint probability of all random variables. This property is rather useful in answering probabilistic queries pertaining to the distributions by which the data is represented in the Bayesian network. The main property used to estimate the network structure is correlation. However recent developments have made it possible to use the more fundamental characteristic of correlation to estimate the networks. Our aim in this project is to explore various ways to construct graph networks based on a certain selection criterion. We begin with total independence between the various Random variables and move to detailing the total dependency among various RVs. In the process we explore the partial representation of dependency and entail the dependency generating sophistication of the algorithm used to encode the partially true network structures. Consequently we develop inference algorithms to determine the mean and entropy of each distribution and relative entropy between distributions (for this purpose we use samples from the original dataset and the synthesized dataset, which in turn is obtained by sampling our networks).

## 2. INTRODUCTION

The data set for this project consists of data drawn from multinomial distribution. They are in the form of samples drawn from a distribution of a set of  $D$  Discrete random variables:

$$x = [X_1, X_2, \dots, X_D]$$

Where the values taken by the variables are  $X_i \in [x_i^0, x_i^1, \dots, x_i^{d_i-1}]$ ,  $i=1, \dots, D$  and  $x_i^j = j$ ,  $j=0, 1, \dots, d_i-1$ . Here  $D=12$  with  $d_i$  ranging from 3 to 5. There are two data sets cursive and handprint. There are also 3 snapshots of the data corresponding to 3 instances of time (2011-2012, 2012-2013, 2013-2014). Each of the dataset contains certain missing values, which we impute using different techniques such as Most likely imputation and k nearest neighbor imputation algorithms.

The first step is to prune the data set of redundant columns which encode tags and other identifiers which are not necessarily important. Here the data generating element or the identifier is the child whose handwriting we examine. This leaves us with 12 features.

The analysis of the handwriting can be useful in many ways. It helps to understand how an individual student's handwriting has changed over the years. It also helps us to understand how the class as a whole is writing in comparison to the Zaner-Bloser copybook style of writing. Here we use both cursive and handprint methods of writing of the word 'and' for comparison. If we

compare the means of students during various years we can find a trend that reflects how their writing style is developing. The entropy of the data points to the individuality of the student in the sense how a student writes different strokes in the word 'and'. We can construct a Bayesian network that can tell us the dependencies of various features extracted on the word 'and' and point to as a whole class how the writing is similar between different students. The Bayesian network can be constructed using many ways. The method we used is the K2 algorithm.

### 3. DATASET

The children's handwriting data, including samples for the same student at different grades, are used for studying the development of handwriting individuality along the time. We study children's handwriting traits' and their developmental trend from the time they begin to learn writing to when the traits are established. The handwriting data are sampled from a large number of students beginning from 2nd grade to 4th grade. Within this time period, the students either begin to learn or just have learned how to write. The collection occurs at every spring, and records how students' handwriting habits develop as handwriting skills continue to mature. Writing samples collected for the same student from different years are examined and studied. The word "and" is chosen because it is one of the few words that children write frequently and repeatedly. After collection, questioned document examiners (QDEs) manually assign different features to handwriting. These are human evaluators who specifically assign the feature vectors to represent the handwriting by 12 features as shown in Figure 1. Ordinal values are assigned to each feature as  $\{0,1,...,k-1\}$  with two exceptions: the value -1 represents missing value and 99 shows inconsistency within the collected sample.






































Initial stroke of "a"	staff right		staff left		staff center			
Formation of "a" staff	tented		retraced		looped		no staff	
Number of "n" arches	one		two					
Shape of "n" arches	pointed		rounded		retraced		combination	
Location of "n" mid	above base		below base		at base			
Formation of "d" staff	tented		retraced		looped			
Formation of "d" initial	overhand		underhand		straight across			
Formation of "d" terminal	curved up		straight		curved down		no obvious end stroke	
Symbol	unusual				symbol			
a-n relationship	a taller		a equal		a smaller			
a-d relationship	a taller		a equal		a smaller			
n-d relationship	n taller		n equal		n smaller			

FIGURE 1(a) 12 feature of cursive data











































# strokes in "a"	one 	two 	three 	uppercase 		
formation of "a" staff	tented 	retraced 	looped 	no staff 	single down 	
# strokes in "n"	one 	two 	three 	uppercase 		
formation of "n" staff	tented 	retraced 	looped 	no staff 	single down 	
shape of arch of "n"	pointed 	rounded 				
# strokes in "d"	one 	two 	three 	uppercase 		
formation of "d" staff	tented 	retraced 	looped 	no staff 	single down 	
initial stroke of "d"	staff top 	bulb 				
Unusual formations	formation 		symbol 			
a-n relationship	a taller 	a equal 	a smaller 			
a-d relationship	a taller 	a equal 	a smaller 			
n-d relationship	n taller 	n equal 	n smaller 			

FIGURE 1(b) 12 feature of handprint data

From the dataset, it is sort of intuitive to group features that are correlated. For example features such as "Initial stroke of a" is dependent on "formation of a stroke", because students who create a stroke in a particular way will most likely form the stroke in the same way. These intuitive groupings give us a better start point while working with Bayesian Networks. This sort of reduces the complexity by reducing the number of combination of nodes required to construct an accurate network.

#### 4. MISSING VALUE IMPUTATION

The dataset contains values that are -1 (missing data) and 99 (inconsistent data). Discarding such data can prove to be very costly. Hence we investigated various missing value imputation technique to fill out these missing data. We clean up the data by assigning "-1" and "99" in the dataset to NaN (not a number). We then applied the k-nearest neighbor imputation to the dataset. The nearest neighbor column is the closest column in Euclidean distance. If the corresponding value from the nearest- neighbor column is also NaN then the next nearest column is used.

#### 5. DATASET VISUALIZATION.

We use various parameters such as mean and entropy of each feature to better visualize and understand the dependencies that will better help us model Bayesian Network. Inference from the mean of the dataset: When the analysis was made on the raw dataset with simplistic metrics

like the mean. Interestingly, the intuition of the gradually improvement in the handwritings of the students was clearly proved to be correct as shown in the following figure. It is clearly indicated that eventually over the years the students improved their handwriting and were more of closer to the reference Zaner-Bloser.

We also observed that the shape of arch of “n” is rounded for handprint as a student progressed from grade 2 to grade 4. This is the similar to Zaner-Bloser. We also observed that the students tend to orient the staff of “a” more towards the right, and tend to have an equal n-d relationship.

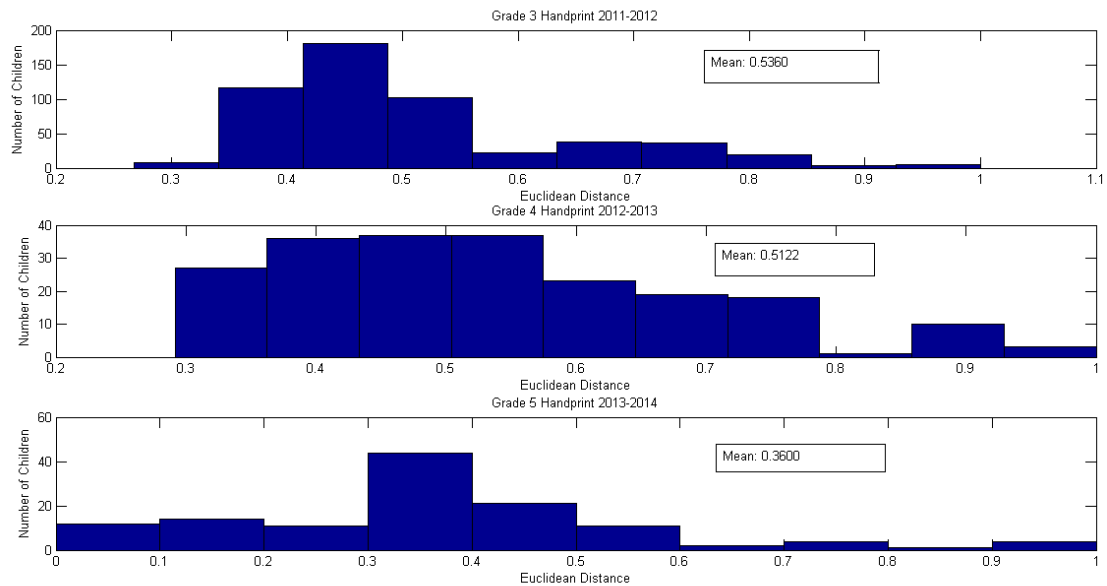
Retracted formation of “d” staff is more prominent in cursive handwriting and the trend of “a” orientation towards the right is also as prominent as in handprint.

Grade 2											
V1		V2		V3		V4		V5		V6	
Min.	:0.0000	Min.	:0.000	Min.	:0.0000	Min.	:0.000	Min.	:0.0000	Min.	:0.0000
1st Qu.	:0.0000	1st Qu.	:1.000	1st Qu.	:0.0000	1st Qu.	:1.000	1st Qu.	:1.0000	1st Qu.	:0.0000
Median	:0.0000	Median	:1.000	Median	:0.0000	Median	:1.000	Median	:1.0000	Median	:0.0000
Mean	:0.2538	Mean	:1.188	Mean	:0.1292	Mean	:1.623	Mean	:0.8997	Mean	:0.2857
3rd Qu.	:0.0000	3rd Qu.	:1.000	3rd Qu.	:0.0000	3rd Qu.	:1.000	3rd Qu.	:1.0000	3rd Qu.	:0.0000
Max.	:4.0000	Max.	:4.000	Max.	:5.0000	Max.	:5.000	Max.	:4.0000	Max.	:4.0000
V7		V8		V9		V10		V11		V12	
Min.	:0.000	Min.	:0.0000	Min.	:0.0000	Min.	:0.0000	Min.	:0.0000	Min.	:0.0000
1st Qu.	:1.000	1st Qu.	:0.0000	1st Qu.	:0.0000	1st Qu.	:0.0000	1st Qu.	:0.0000	1st Qu.	:0.0000
Median	:1.000	Median	:0.0000	Median	:0.0000	Median	:0.0000	Median	:0.0000	Median	:0.0000
Mean	:1.743	Mean	:0.4331	Mean	:0.2386	Mean	:0.3222	Mean	:0.1505	Mean	:0.4985
3rd Qu.	:3.000	3rd Qu.	:1.0000	3rd Qu.	:0.0000	3rd Qu.	:0.0000	3rd Qu.	:0.0000	3rd Qu.	:1.0000
Max.	:3.000	Max.	:3.0000	Max.	:2.0000	Max.	:2.0000	Max.	:2.0000	Max.	:2.0000
Grade 3											
V1		V2		V3		V4		V5		V6	
Min.	:0.0000	Min.	:0.0000	Min.	:0.0000	Min.	:0.000	Min.	:0.0000	Min.	:0.0000
1st Qu.	:0.0000	1st Qu.	:0.0000	1st Qu.	:0.0000	1st Qu.	:1.000	1st Qu.	:0.0000	1st Qu.	:0.0000
Median	:0.0000	Median	:1.0000	Median	:0.0000	Median	:1.000	Median	:1.0000	Median	:0.0000
Mean	:0.6127	Mean	:0.7077	Mean	:0.4401	Mean	:1.208	Mean	:0.6408	Mean	:0.5915
3rd Qu.	:1.0000	3rd Qu.	:1.0000	3rd Qu.	:1.0000	3rd Qu.	:1.000	3rd Qu.	:1.0000	3rd Qu.	:1.0000
Max.	:5.0000	Max.	:4.0000	Max.	:5.0000	Max.	:5.000	Max.	:3.0000	Max.	:5.0000
V7		V8		V9		V10		V11		V12	
Min.	:0.000	Min.	:0.0000	Min.	:0.0000	Min.	:0.0000	Min.	:0.0000	Min.	:0.0000
1st Qu.	:1.000	1st Qu.	:0.0000	1st Qu.	:0.0000	1st Qu.	:0.0000	1st Qu.	:0.0000	1st Qu.	:0.0000
Median	:1.000	Median	:1.0000	Median	:0.0000	Median	:1.0000	Median	:0.0000	Median	:0.0000
Mean	:1.176	Mean	:0.9754	Mean	:0.5176	Mean	:0.6514	Mean	:0.6725	Mean	:0.4261
3rd Qu.	:1.000	3rd Qu.	:2.0000	3rd Qu.	:1.0000	3rd Qu.	:1.0000	3rd Qu.	:1.0000	3rd Qu.	:1.0000
Max.	:4.000	Max.	:2.0000	Max.	:3.0000	Max.	:3.0000	Max.	:3.0000	Max.	:2.0000

**FIGURE 2:** Analysis of Handprint

Since students are being taught the Zaner Bloser style, we assume the Zaner-Bloser style as our benchmark. We then go on to measure the minkowski distance between the individual student's handwriting and the benchmark. This will give us dissimilarity between the Zaner Bloser and the input data.

## 5.1 Similarity



**FIGURE 3:** Handwriting Similarity with Zaner-Bloser

As seen in the graph, the mean value decreases as a student progresses from grade 3 to grade 5 wherein the lower mean score means that the handwriting trend is more similar to the zaner-bloser style.

## 5.2 Entropy

Entropy is a metric to quantify the randomness of a dataset. By analyzing the entropy, we can make conclusion regarding the variation in the behavior of the features as a student advances in their grade.

The entropy of handprint for grade 3 and 4 is as shown below:

Grade	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12
Grade 3	0.42	0.273	0.245	0.344	0.381	0.339	0.291	0.213	0.235	0.337	0.249	0.412
Grade 4	0	0	0.357	0.136	0.2	0.228	0.441	0.228	0.254	0.296	0	0.335

In same grade, between printed and cursive, students tend to be more similar in “Printed”, which has a higher entropy. This is intuitive because, in 3<sup>rd</sup> grade, students are writing cursive for the first time but writing printed for the second time.

## 5.3 Cross-Entropy

KL Divergence is one of the metrics used to evaluate the “Relative Entropy” across the various datasets. This metrics allows us to understand the relationship among the features as how they vary as the students advance in their academics. The lesser the KL divergence value the more strong is the similarity between the distributions of the given feature. The following table shows the KL divergence between the features which were extracted for the same set of students:

2 <sup>nd</sup> Grade-3 <sup>rd</sup> Grade Handprint	5.637	7.595	0.602	15.879	0.797	4.971	2.793	0.12	6.116	2.805	2.351	1.346
3 <sup>rd</sup> Grade Handprint-Cursive	6.241	0.894	2.367	0.616	1.927	0.198	11.948	2.106	0.701	2.289	0.606	0.895
3 <sup>rd</sup> Grade- 4 <sup>th</sup> Grade Handprint	6.313	5.819	0.297	5.049	4.113	6.944	4.371	0.702	0.6	1.494	3.726	2.597

From the above table, we note that few features like 3 and 5 have very low KL divergence. Interestingly, these correspond to *‘Number of  $n$  arches’* and *‘Location of  $n$  mid’*. Thus, as students pass from 2nd grade to 3rd grade, they *tend to write the letter ‘n’ more similarly than other ‘a’ and ‘d’ alphabets* to rest of the class.

KL Divergence between 3rd grade Printed and 3rd grade Cursive indicate that feature 9 is very low. This means the feature set of *“unusual formations” in cursive* and *“Symbols” in printed is almost identical*. This might also be true because of very few such entries.

KL Divergence between 3rd grade Printed and 3rd grade Cursive indicate that *a-d and n-d relationships between cursive and printed is almost identical*, but that between ‘a’ and ‘n’ is not. Overall, in printed writing, the *“number of  $n$  arches” feature is the least divergent across grades*. So, it may be assumed that students write ‘n’ very similarly than they do ‘a’ and ‘d’. This is in sync with what we inferred from the first point.

## 6. BAYESIAN NETWORK

While sometimes BNs are largely constructed manually based on expert knowledge, it is useful to automatically learn their structures from data. Automatic BN structure learning is a very active research area, which contains three main approaches. In actually coming up with our Bayesian network we primarily concentrate on two algorithms:

### 6.1 Graph Construction from correlation

The most basic approach to constructing graph is to use the measure of correlation. We start off by assuming that a node can have at-most two parents. This is basically done to reduce the number of Bayesian structure that we would have to realize otherwise. When we make this assumption, the process becomes less involved. We begin by computing the correlation using the Pearson’s test.

Pearson’s test is a powerful tool that not just gives us an idea about the correlation that exists between two variables but also the directionality. This is useful while deducing features that are dependent and independent. After constructing the graph, the conditional probability distribution can be simply computed by the process of counting. The probabilistic measure is very important when constructing Bayesian networks as this will help us compare the initial reference that we have carried out using the raw dataset and establish the dependencies and independencies between variables. This is also a first step in answering probabilistic query from our Bayesian graph.

### 6.2 Bayesian network construction using the K2 Search Algorithm

The Precise construction of Bayesian network classifier from database is an NP-hard problem. K2 algorithm can reduce search space effectively, to improve learning efficiencies, but it requires the initial node ordering as the input, which is very limited by the absence of priori information.

Let us consider  $B_{Si}$  and  $B_{Sj}$  to be two Bayes networks with identical variables, then

$$\frac{P(B_{Si} | D)}{P(B_{Sj} | D)} = \frac{\frac{P(B_{Si}, D)}{P(D)}}{\frac{P(B_{Sj}, D)}{P(D)}} = \frac{P(B_{Si}, D)}{P(B_{Sj}, D)}$$

From ratios such as above we can rank the structure i.e. parent/child between two nodes. In this algorithm we make four assumptions

1. All the variables/features given are independent
2. Cases occur independently, given a Bayes network model
3. There are no cases that have variables with missing values
4. The density function  $f(B_p | B_s)$  is uniform.  $B_p$  is a vector whose values denotes the conditional-probability assignment associated with structure  $B_s$ .

We then calculate  $P(B_s, D)$

$$P(B_s, D) = P(B_s) \prod_{i=1}^n \prod_{j=1}^{q_i} \frac{(r_i - 1)!}{(N_{ij} + r_i - 1)!} N_{ij} \prod_{k=1}^{r_i} N_{ijk} !$$

Where

D - dataset, it has m cases(records)

Z - a set of n discrete variables:  $(x_1, \dots, x_n)$

$r_i$  - a variable  $x_i$  in Z has  $r_i$  possible value assignment

$B_s$  - a bayes network structure containing just the variables in Z

$\pi_i$  - each variable  $x_i$  in  $B_s$  has a set of parents which we represent with a list of variables  $\pi_i$

$q_i$  - there are  $q_i$  unique instantiations of  $\pi_i$

$w_{ij}$  - denote  $j$ th unique instantiation of  $\pi_i$  relative to D.

$N_{ijk}$  - the number of cases in D in which variable  $x_i$  has the value of  $w_{ij}$  and  $\pi_i$  is instantiated as  $w_{ij}$

$$N_{ij} = \sum_{k=1}^{r_i} N_{ijk}$$

Once, we compute the above probability, we heuristically search where the  $N_{ijk}$  are relative to  $\pi_i$  being the parents of  $x_i$  and relative to a database D. This returns a set of nodes that precede  $x_i$  in the node ordering i.e. the Bayes network.

The ordering required for the K2 algorithm is inferred using the Pearson's test.

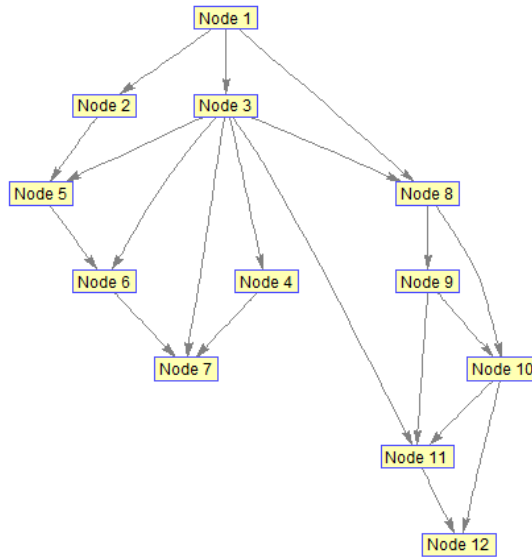


FIGURE 4: BAYESIAN NETWORK FOR HANDPRINT GRADE 3

## 7. INFERENCE

1. From the Bayesian network constructed, we observe that some of our initial intuition about dependencies were indeed true. We observe that Feature 5 (shape of arch of “n”) is dependent on Feature 3 (Stroke of “n”)
2. Number of Stroke in “d” and formation of “d” staff are correlated.
3. Number of stokes in “a” and formation of “a” staff are also dependent.
4. Initial stroke of “d” has a significant effect on “a-d” relationship.
5. We can infer that a-d, n-d and a-n relationships are not dependent on how we write the letters individually but are somewhat only interdependent on each other. This is actually in sync with what we had inferred from the dataset initially, i.e., the prior knowledge that a-d, a-n and n-d relationships might be related to each other because they are about joining two letters.

## 8. BAYESIAN INFERENCE

In a Bayesian Networks we’ll encounter situations where we have some evidence, that is, some of the variables are instantiated, and we want to infer something about the probability distribution of some other variables. The most popular and accurate technique is the exact inference where we analytically compute the conditional probability distribution over the variables of interest. But sometimes that’s too hard to do, in which case we can use approximate techniques based on statistical sampling.

The most usual inference is a conditional probability query. Given the joint distribution over the variables, we can easily answer my question about the value of a single variable by summing (or marginalizing) over the other variables. So, in a domain with four variables, A, B, C, and D, the probability that variable D has value d is the sum over all possible combinations of values of the other three variables of the joint probability of all four values.



We computed the probabilistic queries using the variable elimination method from the Bayesian network that we constructed. The primary and most resourceful probabilistic computation was carried out to understand how many percentage of the total students from a particular grade try to mimic the Zaner Bloser Style. The results are as tabulated below:

QUERY	HANDPRINT	CURSIVE
GRADE 2/ ZANER BLOSER	0.2887	NO DATA AVAILABLE
GRADE 3/ ZANER BLOSER	0.3855	0.3090
GRADE 4/ ZANER BLOSER	0.3090	0.5766
GRADE 5/ ZANER BLOSER	0.2956	0.5422

From the tabulated results we can see that the Grade 3 handwriting for Handprint closely resembles the Zaner Bloser style whereas when it comes to Cursive the same can be observed in Grade 4. There is a significant margin of error here since the variable elimination method is an approximation method used to answer the probabilistic queries.

## 9. MARKOV NETWORK

Bayesian Networks can represent independence constraints that Markov networks cannot. Bayesian Network also takes into account the directionality of the individual parameters that is found using the chi-squared test. Markov Networks come under the undirected graphs and can represent few independence constraints that Bayesian Networks cannot. The challenge here is constructing the Markov Network given the Bayesian Network. This can be done by the process of moralization. For each variable, we add an edge between the variable and its parents and add another edge between all the parents of the variables. A Bayesian networks  $G$  is moral if it contains no immoralities i.e for any pairs of variables  $X, Y$  that share a child, there is a covering edge between two variables.

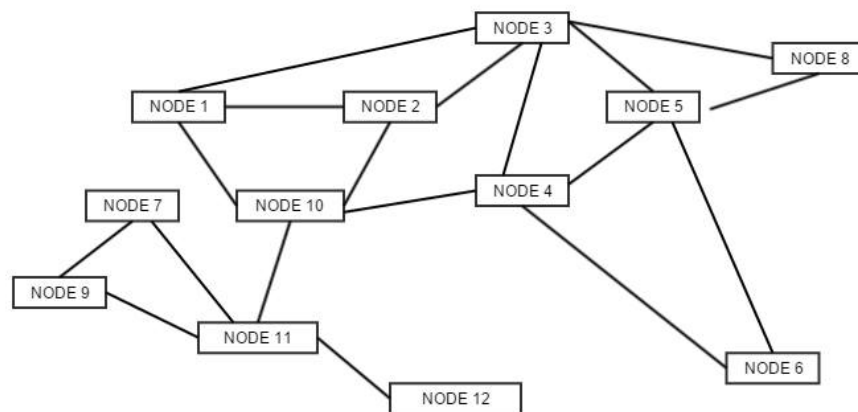


FIGURE 5: MARKOV NETWORK FOR HANDPRINT GRADE 3

## 10. FAILED ATTEMPTS/EXPLORATORY ANALYSIS

In this section, we are going to describe the method which we opted to try out and eventually gain more insight into construction of optimal Bayesian network instead of utilizing the widely used heuristics or greedy methods.

We explored several papers on doing exact structural learning. And finally selected 2 published works [1] and [2]. One was a sequential implementation of doing structural learning and the other was a parallel implementation of the former one.

The parallel implementation didn't materialize within the timeframe of this partial semester because of our lack of expertise in MPI using C++ but this work has immensely motivated us to continue the work and complete it.

### 10.1 Bayesian Structure Learning

The foremost difficulty in constructing a Bayesian network is its super exponential search space in the given number of the random variables. As reported by Robinson, for a set of 'n' random variables there exist possible directed acyclic graphs, where  $r \approx 0.57436$  and  $z \approx 1.4881$ .

$$\text{Number of possible DAG} = \frac{n! \cdot 2^{\frac{n(n-1)}{2}}}{r \cdot z^n}$$

Our work was inspired from the work [1], [2] where the search space was reduced from super exponential space to exponential space. The core idea was to do marginalization over node orders. As for a given graph  $N(X, E)$  of a Bayesian Network. Due to its directed acyclic graph structure, its property is that the parents of a variable precede it in a very specific ordering. By optimizing the scoring function for a given variable and its associated parents will give us the opportunity to discover the optimal Bayesian structure. However, to investigate all possible ordering ( $n!$ -possible) we would run into a complexity of  $O(n!2^n)$  which can be brought down to  $O(2^n)$  by keeping all the parents of variable preceding to a variable irrespective of their ordering.

#### Sequential algorithm:

Scoring function(s): BIC criterion

$F(g, A)$  : Optimal score for a variable 'g' and its predecessor 'A'

$\Pi(A)$  – Ordering of elements from set A

$Q(\Pi(A))$  – Denote the optimal score of a network on A, which is consistent with  $\Pi(A)$

Step 1: Compute  $F(g, \emptyset) = s(g, \emptyset)$  for all  $g \in G$ .

Step 2: For all  $A \subseteq G$ ,  $A \neq \emptyset$  and all  $g \in G$ ,

Compute  $F(g, A)$  as  $\min\{s(g, A), \min_{a \in A} F(g, A - \{a\})\}$ .

Step 3: Set  $M(\emptyset) = \emptyset$ .

Step 4: For all  $A \subseteq G$ ,  $A \neq \emptyset$ , do the following two steps:

Step 4a: Compute  $g^* = \arg \min_{g \in A} (F(g, A - \{g\}) + Q^{A - \{g\}}(M(A - \{g\})))$ .

Step 4b: For all  $1 \leq i < |A|$ , set  $M(A)(i) = M(A - \{g^*\})(i)$ , and  $M(A)(|A|) = g^*$ .

Step 5: return  $Q^G(M(G))$ .

**Complexity analysis:** The above algorithm takes the advantage of the dynamic programming strategy which exploits the optimal substructure property.

Step 1 & step 2:  $O(n \cdot 2^n)$

Step 3 & Step 4:  $O(n \cdot 2^n)$

Overall Complexity:  $O(n \cdot 2^n)$

Parallel version of the above algorithm: Which brings the computational complexity to  $O(\frac{1}{n} 2^n)$  at no extra space complexity than the sequential algorithm.

The above algorithm can be visualized as dynamic programming strategy applied on the lattice structure formed by the partial order of the power set of  $X$ . In this work [cite zola and thesis], the lattice  $L$  is viewed as a directed graph  $(V, E)$ , where  $V = 2^X$ , and  $(B, A) \in E$  if  $B \subset A$  and  $|A| = |B| + 1$ . And the lattice is naturally divided into levels where each level  $l$  contains all the subsets of size  $l$ . And the algorithm was formalized by mapping the nodes to processors and the having edges represent communication if the incident nodes are assigned to different processors. A node  $A$  at level  $l$  has  $l$  incoming edges from nodes  $A - \{X_j\}$  for each  $X_j \in A$ , and  $n - l$  outgoing edges to nodes  $A \cup \{X_i\}$  for each  $X_i \notin A$ . Functions  $Q^*(A)$ ,  $\pi^*(A)$ , and a total of  $(n - l)$   $F$  functions are computed at node  $A$ . All of these values need to be sent along each of the outgoing edges. On an outgoing edge to node  $A \cup \{X_i\}$ , the  $F(X_i, A)$  value is used in computing  $Q^*(A \cup \{X_i\})$ , and the remaining  $F(X_k, A)$  ( $X_k \notin A$  and  $X_k \neq X_i$ ) values are used in computing  $F(X_k, A \cup \{X_i\})$  values at node  $A \cup \{X_i\}$ . Note that each of the  $(n - l)$   $F$  values at  $A$  are used in computing the  $Q^*$  value at one of the ' $n - l$ ' nodes connected to  $A$  by outgoing edges. Each level in the lattice can be computed concurrently, with data flowing from one level to the next.

## 11. CONCLUSION

Using the Bayesian network construction, we find out that most students orient the letter 'a' towards the right which becomes more prominent in cursive handwriting. We also note that in handprint, no student ignores the obvious end stroke of the letter 'd' after the second grade. This is noted from the fact that no sample has a feature 8 value more than 2 after the third grade. In same grade, between printed and cursive, students tend to be more similar in "Printed", which has a higher entropy. This is intuitive because, in 3rd grade, students are writing cursive for the first time but writing printed for the second time. We also observe from the similarity test with the Zaner-Bloser style that as a student progresses, their handwriting becomes more similar to the Zaner-Bloser style. In the future, we plan to extend this to a parallel Bayesian structure realization to obtain even better and accurate results.

## 12. References

- [1] S. Ott, S. Imoto, and S. Miyano, "Finding optimal models for small gene networks.," in *Pacific Symposium on Biocomputing (PSB)*, 2004, pp. 557–567.
- [2] B. O’Gorman and a Perdomo-Ortiz, "Bayesian Network Structure Learning Using Quantum Annealing," *arXiv Prepr. arXiv ...*, pp. 3546–3551, 2014.
- [3] O. Nikolova, "Parallel Algorithms for Bayesian Networks Structure Learning with Applications to Systems Biology," *2011 IEEE Int. Symp. Parallel Distrib. Process. Work. Phd Forum*, pp. 2045–2048, 2011.

- [4] M. Puri, S. N. Srihari, and Y. Tang, "Bayesian network structure learning and inference methods for handwriting," *Proc. Int. Conf. Doc. Anal. Recognition, ICDAR*, pp. 1320–1324, 2013.
- [5] Y. Tang and S. N. Srihari, "Efficient and Accurate Learning of Bayesian Networks using Chi-Squared Independence Tests," *Int. Conf. Pattern Recognit.*, no. Icpr, pp. 2723–2726, 2012.