An Empirical Validation of Recursive Noisy OR (RNOR) Rule for Asthma Prediction

Vibha Anand[1, 2], M.S., Stephen M. Downs[1, 2], M.D., M.S.

[1] *Children's Health Services Research, Indiana University School of Medicine*
[2] *Regenstrief Institute for Health Care, Indianapolis, IN*

Abstract

In 2004, an extension of the Noisy-OR formalism termed the Recursive Noisy–OR (RNOR) rule was published for estimating complex probabilistic interactions in a Bayesian Network (BN). The RNOR rule presents an algorithm to construct a complete conditional probability distribution (CPD) of a node while allowing domain causal relationships over and above causal independence to be tractably captured in a semantically meaningful way. However, to the best of our knowledge, the accuracy of this rule has not been tested empirically. In this paper, we report the results of a study that compares the performance of a data-trained expert BN (empiric BN) with the reformulated BN, using the RNOR rule. The original empiric BN was trained with a large dataset from the Regenstrief Medical Record System (RMRS). Furthermore, we evaluate conditions in our dataset which render the RNOR rule inapplicable and discuss our use of Noisy-OR calculations in such situations. We call this approach "Adaptive Recursive Noisy-OR".

Keywords:
Asthma; Bayesian Network, Noisy-OR, Recursive Noisy OR rule (RNOR), Adaptive Recursive Noisy OR

Introduction

Bayesian belief networks are a well established formalism for representing conditional probability dependencies and independencies among random variables. [1] They represent an effective method for supporting reasoning under uncertainty in decision support systems.[2] Algorithms have been developed for training belief networks by learning their conditional probability distributions (CPD) from datasets.[3-4] However, a challenge arises when the training data set has no cases representing a particular combination of values for variables that condition a particular CPD. This is a common problem in complex belief networks even when large training sets are available. [5] A common strategy in this situation is to assign a uniform (uninformed) distribution to the dependent variable, conditioned by this combination of variables. For example, when the probability of asthma is conditioned on the sex, race, insurance and past wheezing history of a patient, there may be no cases in the training data that are male, white, on Medicaid and with a positive history of wheezing. Under the uniform distribution strategy, the probability of asthma would have a 50-50 distribution.

An alternative approach is to assume causal independence among predecessors (parents) of a given node. In this case, it may be reasonable to apply a Noisy-OR calculation [1] to estimate the probability of the child node given a particular combination of values for the parents. In 2008, we compared a reduced network based on this Noisy-OR formulation to the same network trained with the data alone, and we showed the noisy-formulation performed almost as well in predicting outcomes in two different test sets of data.[6]

However, the ideal strategy would retain posterior distributions for combinations of parent node values that exist in the training set while applying the Noisy-OR rule when there are no cases representing a combination of conditioning variables. In 2004, a potential solution to this problem was published by Lemmer and Gossink. [7] The Recursive Noisy-OR (RNOR) rule described by these authors was intended to incorporate expert estimates of probabilities conditioned on more than one node while applying the Noisy-OR rule when these higher order conditional probabilities were not available. We reasoned that the RNOR algorithm might be a successful strategy for training a belief network from a data set that did not contain cases representing all combinations of variables conditioned on a given node. This paper describes the development and evaluation of this strategy.

Methods

We constructed a BN in the domain of asthma prediction in children, using expert knowledge to derive a directed acyclic graph (DAG) and applied a commercially available software package to "learn" the CPDs from a large clinical dataset. This empiric BN has been described elsewhere. [8] For this study, we reformulated the CPDs in our domain expert's BN using the RNOR rule. Our empiric BN and RNOR BN were tested against two independent clinical data sets. The first (RMRS dataset) consisted of 1/3 of the dataset used to train the empiric BN that was randomly split from the data before training. The second dataset was collected prospectively and derived from our clinical decision support system that uses RMRS data but makes an independent assessment of asthma diagnosis.

EMR Data and Variables

Data – Clinical data for this study were derived from two datasets. The first consisted of 16,187 cases from the Regenstrief Medical Record System (RMRS). RMRS is an inpatient and outpatient information system which contains over 30 years of data and more than 300 million numeric or coded patient observations.[9] Data for this study had been identified in a separate large retrospective cohort study to identify variables measured before age 2 that predict an outcome of asthma after age 5. The data collected were for children ages ≥ 5 years.

The second dataset was derived prospectively from the CHICA (Child Health Improvement through Computer Automation) decision support system. [10] CHICA receives data from RMRS as part of the patient's visit but assesses the presence of asthma independently. Table 1 lists our model variables. At the time of the study, the data pulled from the CHICA system consisted of 1982 cases. The baseline characteristics of the training set and the test sets are reported in our previous study. [6]

Bayesian Network and Noisy-OR model: The Noisy-OR gate (model) is a member of the family of models referred to as independence of causal influences (ICI). [4, 11-12] A leaky Noisy–OR model can be formally described as below for several possible causes $x_1, x_2 \ldots x_n$ of an effect y using two assumptions - (a) each of the causes x_i has a probability (p_i) of producing the effect in the absence of all other causes, and (b) each cause is sufficiently independent of the presence of other causes. The probability p_i is also called the *link* probability and determines the causal strength in the absence of all other causes. It has been shown [1] that the probability of *y* given a subset X_p (x_i present) is given by the following equation –

$$P(y \mid X_p) = 1 - \prod_{x_i \in X_p} (1 - p_i) \qquad (1)$$

From [1] it has also been shown that the link probability is related to the global leak probability p_0 and the probability p' which is the link probability in absence of all causes including un-modeled causes (i.e. leak) by the following equation

$$p_i = P(y \mid x_{i\ only}) = p' + p_0 - p' * p_0 \qquad (2)$$

Calculating p_i or p' for computing Noisy-OR is attributed to two different ways of parameterization of CPD credited to Henrion [13] and Diez [5] respectively. For our study we used the Henrion parameter (p_i) to derive the Diez parameter (p') using equation (2).

Variable	Values
Race *	White, Black, Hispanic, Other, Unknown
Sex *	Male, Female
Eczema	True, False
Wheeze	ICD9 or clinic billing diagnosis before age 2 (True, False)
Asthma	ICD9 (493.*) or any clinic billing diagnosis after age 5 or at least 3 drugs from a specified list within 12 months after age 5 (True, False)
X-ray	Chest x-ray before age 2 (True, False)
Drug	Drugs from specified list (True, False)
Wz_hosp	Inpatient admission with hospital ICD9 as wheezing (True, False)
Wz_er	Any ER visit with billing ICD9 as wheezing (True, False)
Ins_cat *	Insurance category - first available insurance in the same year of the first wheezing diagnosis – Medicaid., Private, Selfpay, Unknown

Table 1 – Data Variables for Model (* Denotes variables are not Boolean)

Bayesian Network and Recursive Noisy-OR (RNOR) model: In 2004, a rule for estimating complex probabilistic interactions was published by Lemmer and Gossink [7]. The rule builds upon the Noisy OR equation (1) to accommodate non-independent causes of an effect for calculating RNOR probability $p^R(x)$. Interested readers are encouraged to refer to the original paper [7]. The RNOR rule is a generalization of Noisy-OR and reduces to Noisy-OR in cases where a subset of a single cause is provided by the expert $p^E(x)$. Furthermore, the rule preserves certain ratios (synergies and interference, see below), and the authors claim a major advantage of the rule is that it allows for arbitrary causally dependent subsets of probability scores to be incorporated in the estimation of $p^R(x)$, the probability score of effect.

The rule states that as long as the expert provided values for subset of causes do not imply inhibition (i.e. abides by positive causality; see synergy and interference) it can be applied. Expert provided values implying inhibition among a set of causes will cause RNOR to produce probability scores greater than one, rendering the rule inapplicable. The RNOR rule is summarized in the following equation-

$$P^R(x) = \begin{cases} p^E(x), \text{ if expert provides info} \\ 1 - \prod_{i=0}^{n-1} \dfrac{(1 - p^R(x \setminus \{x_i\}))}{(1 - p^R(x \setminus \{x_i, x_{(i+1) \bmod n}\}))}, \text{ otherwise} \end{cases} \qquad (3)$$

Synergy and Interference: To evaluate whether the numbers computed to results are semantically meaningful [7] the authors introduce the notion of positive causality. *Positive causality* refers to the idea that additional causes always increase the probability of achieving an effect. *Synergy* of causes satisfies positive causality and produces an effect (probability) greater than the Noisy-OR calculation. *Interference* of causes also satisfies positive causality but produces an effect less than the Noisy-OR calculation. However, inhibition, in contrast to interference, violates positive causality and will produce probability scores greater than one. Therefore, the "information on the probability of an effect from a combination of causes provided by an expert (or derived from data) can be represented as a scalar multiple of the regular Noisy-OR." This can be represented using equation (4) as defined in the original paper. [7]

$$1 - p^{R}(x) \equiv \delta(x)(1 - p^{N}(x)) \qquad (4)$$

where $p^R(x)$ represents the probability from a RNOR estimation and $p^N(x)$ represents the standard Noisy-OR estimation. The factor $\delta(x)$ represents a scalar gain or attenuation coefficient between the two estimates. If $\delta(x)$ is less than one then it represents a biased amplifying coefficient of the probability and hence *synergy*. Conversely if $\delta(x)$ is greater than one then it represents a biased attenuating coefficient and hence *interference*. Finally, if it is equal to one then an independent product combination could hold implying causal independence

Asthma Prediction BN using data derived CPD: As with the previous study [6], we constructed a BN for asthma prediction, using our medical expert's (SMD) domain knowledge of clinical variables (Figure 1, Table 1) and Netica software (www.norsys.com). [14] Netica allows Bayesian network construction and parameter learning from data (Empiric BN). We derived the parameters for CPD construction with training data from RMRS.

Deriving CPD from Data: The data from the 16,187 cases in the RMRS were randomly split into ~2/3 of cases (11,000) for a training set and ~1/3 (5,187) for a test set. This test set formed our empiric dataset. Using the predictor data variables from Table 1 as nodes and the domain knowledge for joining them with arcs, the domain expert (SMD) who is both trained in pediatrics and in information sciences created a DAG as shown in Figure 1. An 18.6% marginal probability of asthma was learnt from data for this DAG. On examining the CPD of asthma, we found the majority of its rows contained uninformed priors (i.e., uniformly distributed probability score - asthma=50-50). We attributed this to the lack of cases satisfying the particular combination of causes in the training set (e.g., race = Other, Sex = male, Ins_cat = selfpay, Eczema = true, Wz_hosp = true). We sought to use the RNOR rule to calculate conditional probability for the CPD row value for asthma in such cases.

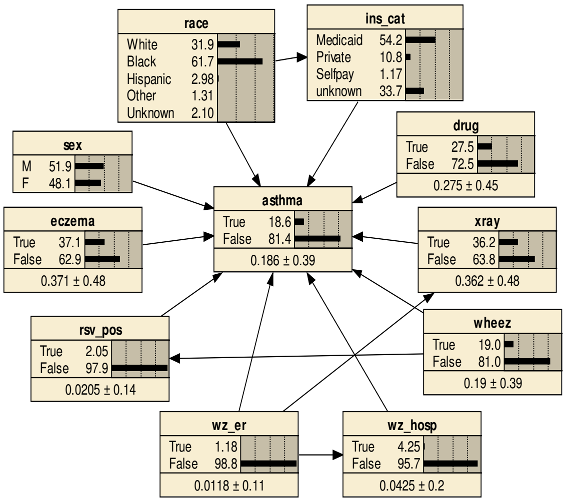

Figure 1: BN with CPD

Asthma Prediction BN using RNOR Rule CPD: Using expert's BN from Figure 1, we calculated the CPD of the asthma node using the RNOR rule. The Noisy-OR model and therefore the RNOR model inherently assumes binary causes (absent/present; true/false); therefore we dichotomized the non-binary nodes i.e. *race and, insurance category* (marked with * in table 1 & 2) into "true" and "false" condition by assigning "true" to the state that minimized the global leak (p_0) when all the other nodes are in a "false" state. Thus, because boys were more likely to have asthma, male sex was coded as the "true" state.

Deriving link probabilities for RNOR rule: To apply the RNOR rule to the network for CPD computation of asthma, we calculated the link probabilities (p') of each cause of the effect (asthma) individually (Table 2).

P(eczema=True)=0.048	P(drug= True) = 0.117
P(rsv_pos=True) = 0.4	P(wheez=True) = 0.17
P(wz_er=True) = 0.37	*P(ins_cat=Private)= 0.058
P(wz_hosp=True)=0.104	*P(race=Black)=0.042
P(xray=True) = 0.072	P(sex=Male) = 0.022

Table 2 – Link probability of each node to asthma
(*denotes "true" state for these variables)

This was a two step process – (a) reduce the network by absorbing all nodes (see below node absorption) except the node in question (e.g. eczema), looking up the "true" conditional probability score given the cause (i.e. asthma=True given eczema=True), then looking up its "false" conditional probability score to find leak p_0 for this reduced network and (b) using equation (2) to find the probability (p') in absence of all un-modeled causes including the leak calculated in (a). We repeated this two step process for all causal nodes for asthma thus first deriving the Henrion parameter (p_i) which includes the leak parameter (p_0) in the posterior probability. From there, we derived the link probability p´ (Diez parameter) for each cause alone.

Node absorption is a network transform which removes nodes from a BN and makes any necessary adjustments to the resulting network. Also known as averaging out or "summing out a variable", this transform leaves the full joint probability distribution of the remaining nodes unchanged.

Applying RNOR rule using link probabilities: Using the Java programming language and the Netica application programming interface, we programmed the RNOR algorithm as defined in [7]. The algorithm calculated the probability score of asthma in each row of the CPD using equation (3). An example combination - (s)ex = Male, (r)ace = Private, (w)heez = True is given here -

$$P^R(x) = 1 - \frac{1 - P(r,\ w)}{1 - P(w)} \times \frac{1 - P(w,\ s)}{1 - P(s)} \times \frac{1 - P(s,\ r)}{1 - P(r)} \quad (5)$$

In the example above (equation 5), the RNOR algorithm calculated the row value for the CPD by first calculating each of the subsets, multiplying them and subtracting from 1. The probability score for asthma for each combination of conditioning variables was calculated in three different ways–

Using RNOR rule: The algorithm recursively calculated the probability score for successive combinations; <u>using the original scores, $p^E(x)$, from data learnt CPD wherever they existed (i.e., not 50-50)</u>, using equation (3). For example, if P(Wheez, Sex) was non-uniform (i.e., not 50-50) in the original data learnt CPD of asthma (from the empiric BN), it was used in successive calculations for higher order combinations. When the resulting probability was negative or exceeded 1, suggesting inhibition, the CPD row was left with 50-50.

Noisy-OR approach: Starting with link probabilities listed in table 2, the algorithm recursively calculated a probability score of asthma for successive combinations of conditioning variables <u>without using any values from the original data learnt CPD</u> from the empiric BN. This essentially reduced the model to a *Noisy-OR* model.

Adaptive Noisy-OR approach: As with the RNOR calculations, the algorithm computed CPD values recursively using $p^E(x)$ where applicable. However, for those scores resulting in negative values, instead of leaving a uniform distribution, we used the Noisy-OR value, reasoning that it would be a better approximation than the 50-50 value in the CPD row. We termed this "Adaptive Recursive Noisy-OR" or Adaptive RNOR (ARNOR).

BN with CPD calculated from these three methods were used in two separate tests, comparing their predictive abilities in the two test sets.

Testing the Models: The BN models were evaluated, first, using data from our original test set of RMRS data (1/3 split from the large cohort study 5187 cases) and, second, using the data set derived from the prospectively collected CHICA database (1982 cases), utilizing Netica's test interface. We compared the sensitivity, specificity, positive predictive value, and negative predictive value using Receiver Operating Characteristic (ROC) curves [15]. The ROC curve was obtained by plotting pairs of true positive rate (sensitivity) and false positive rate (1 - specificity). The area under the ROC curve (AUC) was used as a measure of overall test performance. AUCs were compared using the method of Hanley and McNeil [16] which specifically accounts for comparing ROC derived from same cases.

Results

We had 5187 cases in our RMRS test set and 1984 cases in the independent test set (CHICA). Results for the tests are shown in Figures 3 and 4 respectively.

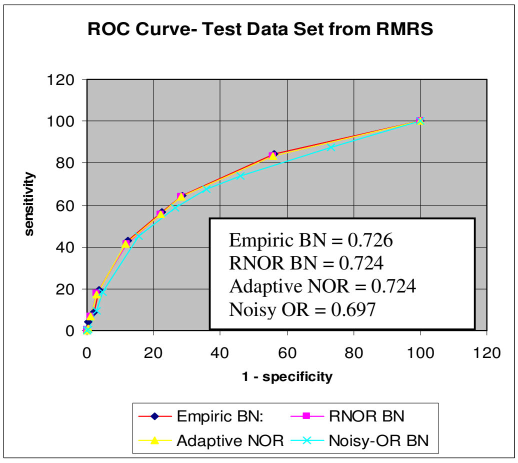

Figure 3 – Evaluation using test data from RMRS

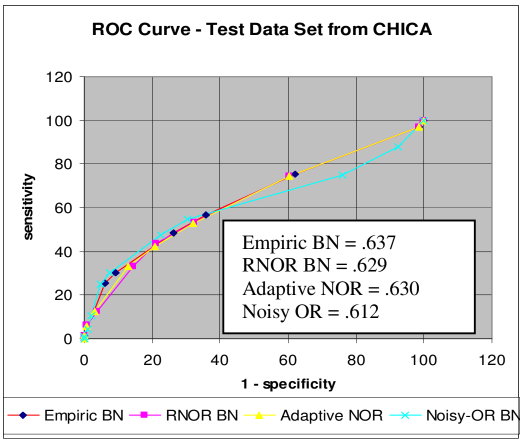

Figure 4 – Evaluation using test data from CHICA

There were no statistically significant differences between the predictive ability of the RNOR or Adaptive RNOR (ARNOR) and that of the empirically trained BN. Both RNOR and Adaptive RNOR had larger AUC than the Noisy-OR BN in both RMRS and CHICA datasets but the difference did not reach statistical significance (Table 3).

Dataset	AUC1	AUC2	P-value
RMRS	Empiric BN	RNOR	0.956
RMRS	Empiric BN	ARNOR	0.968
CHICA	Empiric BN	RNOR	0.620
CHICA	Empiric BN	ARNOR	0.625
RMRS	Noisy-OR	RNOR	~1.0
RMRS	Noisy-OR	ARNOR	~1.0
CHICA	Noisy-OR	RNOR	~1.0
CHICA	Noisy-OR	ARNOR	0.851

Table 3 – AUC Comparisons using Hanley and McNeil method for same cases

Discussion

We anticipated that the RNOR would perform better than the empirically trained BN. The RNOR retains the higher level posterior distributions that can be extracted from the data, but in the absence of cases in the data, the RNOR estimates posterior distributions that are more reasonable than a uniform distribution. This is especially true for a variable like *asthma* in this dataset where the posterior probability would not be expected to reach 50% as the data variables measured were before age 2 to predict an outcome of asthma at age 5 or later. We observed such a trend, but the differences were not statistically significant.

In our application of the RNOR we ignored any negative probability scores that are produced (as a result of recursive calculations in the algorithm) and believe they are analogous to the interaction term in a regression setting. We believe these negative scores are produced where inhibition exists between subsets of dependent causes and the scores produced by the algorithm are greater than one. One such example is – *sex=Male, drug=True, xray=True* – inhibition between *drug* and *xray* (subset score=1.0034) and *sex* and *drug* (score=1.03), though *xray* and *sex* subset had a score within bound (score=0.969). The implication is that the domain chosen for this study may not be an ideal application of the RNOR strategy but the RNOR algorithm can be used to detect inhibition in large datasets. Nonetheless, the RNOR had no less predictive value than the BN derived directly from data.

With our Adaptive Recursive Noisy-OR approach; using Noisy-OR calculated value to score variable combinations with inhibition, we produced CPD values within bound and with informed priors that resulted in a similar predictive value as the RNOR and empiric BN. Therefore, we are optimistic that RNOR and the Adaptive RNOR approach holds promise to improve the predictive power of empirically trained belief networks.

References

1. Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. 1988.
2. Middleton, B., et al., *Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base. II. Evaluation of diagnostic performance.* Methods Inf Med, 1991. **30**(4): p. 256-67.
3. Heckerman, D.E., Geiger, D, Chickering, *Learning Bayesian Networks: The combination of knowledge and statistical data.* Machine Learning, 1995. **20**: p. 197-243.
4. Heckerman, D.E., Breese, J.S, *Causal Independence for Probability Assessment and Inference Using Bayesian Networks.* IEEE Transactions on Systems, Man and Cybernetics, 1996. **26**: p. 826-831.
5. Diez, F., *Parameter adjustment in Bayes networks. The generalized noisy-OR-gate.* Proceedings of the 9th Conference on Uncertainity in Artifical Intelligence (UAI-94), 1993: p. 286-292.
6. Anand, V. and S.M. Downs, *Probabilistic asthma case finding: a noisy or reformulation.* AMIA Annu Symp Proc, 2008. **6**: p. 6-10.
7. Lemmer, J., Gossink DE, *Recursive Noisy OR - A Rule for Estimating Complex Probabilistic Interactions.* IEEE Transactions on Systems, Man, And Cybernetics - Part B: Cybernetics, 2004. **34**(6).
8. Anand, V. and S.M. Downs, *Probabilistic Asthma Case Finding - A Pilot Study using the CHICA system.* Medinfo, 2007. **12**(Pt 1): p. 292.
9. McDonald, C.J. and et al, *The Regenstrief Medical Record System: 20 years of experience in hospitals, clinics, and neighborhood health centers. MD Computing.* 1992. **9**(4): p. 206-17.
10. Anand, V., et al., *Child Health Improvement through Computer Automation: the CHICA system.* Medinfo, 2004. **11**(Pt 1): p. 187-91.
11. Heckerman, D.E., Breese, J.S, *A new look at causal independence.* Proceedings of the 10th Annual Conference on Uncertainity in Artifical Intelligence (UAI-94), 1994: p. 286-292.
12. Srinivas, S., *A Generalization of the Noisy-Or Model.* Proceedings of the Ninth Annual Conference on Uncertainty in Artificial Intelligence (UAI-93), 1993: p. 208-215.
13. Henrion, M., *Some practical issues in constructing belief networks.* Uncertainity in Artificial Intelligence 3. New York, NY: ELsivier Science Publishing Company, Inc., 1989: p. 161-173.
14. *Netica. Application for Belief Networks and Influence Diagrams. Norsys Software Corp. 1997.*
15. Hanley, J.A. and B.J. McNeil, *The meaning and use of the area under a receiver operating characteristic (ROC) curve.* Radiology, 1982. **143**(1): p. 29-36.
16. Hanley, J.A. and B.J. McNeil, *A method of comparing the areas under receiver operating characteristic curves derived from the same cases.* Radiology, 1983. **148**(3): p. 839-43.