Bayesian Networks' Development Based on Noisy-MAX Nodes for Modeling Investment Processes in Transport

Volodymyr Lytvynenko[1[0000-0002-1536-5542]], Nataliia Savina[2[0000-0001-8339-1219]], Jan Krejci[3 [0000-0003-4365-5413]], Mariia Voronenko[1[0000-0002-5392-5125]], Maryna Yakobchuk[2[0000-0003-1587-1370]], Olena Kryvoruchko[2[000-0002-5426-7347]]

[1]Kherson National Technical University, Kherson, Ukraine,
[2]National University of Water and Environmental Engineering, Rivne, Ukraine,
[3]Jan Evangelista Purkyne University in Usti nad Labem, Czech Republic,
immun56@gmail.com, mary_voronenko@i.ua,
n.b.savina@nuwm.edu.ua, 2marysa2@ukr.net,
o.p.kryvoruchko@nuwm.edu.ua,
jan.krejci@ujep.cz

Abstract. This article focuses on the use of Bayesian networks for analyzing the growth relationship of Ukraine's gross domestic product (GDP) from the volume of investment in the transport industry and offers a comparative description of the use of different structural training algorithms. It is shown that Noisy-max nodes as compared to General nodes provide relatively high initial accuracy. General nodes require a repeated validation procedure. When using the Hirerical sampling method, the accuracy of the network result with General nodes remains unchanged, and with Noisy-max nodes, it increases (in our case, by 12.32%). However, Noisy-max nodes entail an increase in time and computational costs.

Keywords: Transport industry; General nodes; Noisy-max nodes; Bayesian networks; Structural learning; Parametric learning; Sensitivity analysis; Validation.

1 Introduction

Successful implementation of the investment policy will contribute to the implementation of one of the main country's economy tasks to increasing the number of main domestic investment resources sources. This will create the necessary prerequisites for the production growth and expanded reproduction of GDP in order to increase the population's well-being.

Therefore, it would be advisable to determine the informative investment indicators that have the greatest impact on the dynamics of Ukraine's GDP. It is necessary to develop a model of the relationship between capital investment and GDP.

In [1], the author defined investment as “the current increase in capital property values as a result of production activities during a given period,” or as a share of income for a given period that was not used for consumption.

It is believed that investment is the material basis for the economy's structuring. Solving the problem of investment will mean the beginning of not only the economy's restructuring but also its stabilization and subsequent growth [2].

The term "investment" has several meanings. First, it means the purchase of shares, bonds with the expectation of certain financial results. Secondly, real assets, for example, machinery, equipment necessary for the production and sale of any goods. In the broadest sense, investments provide the mechanism necessary to finance the growth and development of the country's economy, region, industry or enterprise.

Successful implementation of the investment policy will contribute to the implementation of one of the main tasks of the country's economy to increasing the number of main sources of domestic investment resources. This will create the necessary prerequisites for the production growth and expanded reproduction of GDP to increase the population well-being.

An important problem is the identification of causal relationships between causes or factors affecting GDP. Therefore, it would be advisable to determine informative investment indicators that have the greatest impact on the dynamics of Ukraine's GDP. The results presented in the article concern research on the development of probabilistic deterministic models using Bayesian networks to identify the factors affecting investment in the transport industry of Ukraine’s GDP.

Considering that one of the difficulties in the development of Bayesian networks is an exponential increase in the parameters number in their conditional probability tables (CPT), this study proposes a technique for using noisy-MAX nodes to simulate economic processes.

The purpose of the study is constructing a Bayesian network model using noisy-MAX nodes for analyzing the dependence of Ukraine's GDP growth on the volume of investments in the transport industry.

2 Problem Statement

For a set of events $X^{(i)}, i=1,\dots,N$ that are related, and a set of learning data $D=\left(d_1,\dots,d_n\right), d_i=\left\{x_i^{(1)}x_i^{(2)}\dots x_i^{(N)}\right\}$, is given. Here the subscript is the observation amount, and the upper one is the variable amount, n –is the amount of observations, each observation consists of $N\left(N\geq 2\right)$ variables, and each j -th variable $\left(j=1,\dots,N\right)$ has $A^{(j)}=\left\{0,1,\dots,\alpha^{(j)}-1\right\}$ $\left(\alpha^{(j)}\geq 2\right)$ conditions. Based on a given training sample, you need to build an acyclic graph connecting the event sets $X_i, i=1,\dots,N$. In addition, each BN structure $g\in G$ is represented by a set N of predecessors $\left(P^{(1)},\dots,P^{(N)}\right)$, that is, for each vertex $j=1,\dots,N, P^{(j)}$ it is a variety of parent vertices, such that $P^{(j)} \subseteq \left\{X^{(1)},\ldots,X^{(N)}\right\} \backslash \left\{X^{(j)}\right\}$. In this study, the Bayesian network of modeling investment processes in transport and their impact on GDP are built using noisy-MAX nodes. To do this, we have events $X^{(i)}, i=1,\ldots,N$ that are affected by the uncertainties of a different nature. And also we have data describing these events.

3 Review of the Literature

There is a wide variety of Bayesian networks applications, including design [3], consumer behavior [4], social behavior [4], support for clinical decision making [5, 6], system biology [6], ecology [7], and so on.

The use of Bayesian networks in socio-economic research is widely considered in [8,9], where they are one of the mathematical tools for analyzing social behavior, since they allow describing, modeling and predicting any empirical data: quantitative, qualitative, and also data of mixed nature. Bayesian networks make it possible to use both probabilities obtained by analytical or statistical methods and expert estimates, as shown in [5, 6, 7].

The advantage of using Bayesian networks is their resistance to incomplete, inaccurate and noisy information. In these cases, the result will reflect the most likely outcome of events [8, 10].

One of the main problems of Bayesian networks is the rapid increase in parameters while increasing the number of parents. To solve this problem, the most widely used are Noisy-MAX [11, 12], since they use multivalued variables. This approach has proven itself in many real-world applications [13, 14, 15]). A small amount of parameters, that will be enough to indicate the entire CPT is a major advantage. This allows improving the quality of the distributions extracted from the data [16], as well as reducing the spatial and temporal complexity of the algorithms for the BN [8,17].

4 Materials and Methods

A pair <***G***, *B*> called a Bayesian network (BN), when the first part of G is a acyclic directed graph corresponding to random variables. When each variable is autonomous of its parents in G, so a graph is written as a composition of autonomous conditions. The second part of the pair, B, is the composition of parameters defining the network. It composed of parameters $\mathrm{Q}_{x^i|pa(X^i)} = P(x^i \mid pa(\mathrm{X}^i))$ for each possible x^i value from X^i and $pa(X^i)$ from $Pa(X^i)$, where $Pa(X^i)$ means the variable X^i parents set in G. Each variable X^i in graph G is suggested as a vertex. If we consider more than only one graph, then we use the notation to identify the parents $Pa^G(X^i)$ in graph G.

The BN's cumulative probability B is determined by the equation $P_B(X^1,...,\mathrm{X}^N) = \prod_{i=1}^{N} P_B(\mathrm{X}^i \mid \mathrm{Pa}(\mathrm{X}^i))$.

The BN represents a model for getting probabilistic dependencies, as well as the absence of these dependencies. At the same time, the A→B relationship can be causal, when event A causes B to occur. So that is, at the time, when there is a mechanism by which value is accepted by A affects the value adopted by B. When all BN's connections are causal, so BN is called causal.

From the existing discretization methods (hierarchical discretization, discretization on the same width of classes, discretization on the same number of points inside the classes) for the existing data set a hierarchical discretization method was chosen [18,19.].

Structural methods of BN training are algorithms, such as: the Bayesian Search, the Essential Graph Search, the TAN. In our study, the Greedy Thick Thinning algorithm is used.

The essence of the Greedy algorithm for constructing the BN structure is as follows. The structure learning algorithm, called Greedy Thick Thinning (GTT), is found on the approach of Bayesian search. GTT was proposed in [6]. The GTT algorithm begins with the construction of an empty graph and then the stepwise multiple additions of an arc. This process occurs without creating a cycle.

Arcs are added to maximize the marginal likelihood P (D | S). This process is repeated until the addition of the arc leads to a positive increase. This phase is called "thickening".

Then the arcs are removed step by step until the removal of the arc leads to a positive increase in P (D | S). This phase is called "thinning". The algorithm is quite effective due to the fact that it is exposed to the trap of local maxima. In GTT, you can use two priorities. The priority of *BDeu* provides an equal score through the equivalent of Marcov's classes. The priority of *K2* is constant in all variables and, as a rule, is used to maximize *P(G|D)* when directly searching for a space of graphs.

Validation of the developed network was carried out according to the algorithm of maximizing the expectation, which was proposed for the first time in 1977 in [20]. The algorithm finds local optimal estimates of the maximum likelihood of parameters. If the values of all nodes are known, then training (at some step M) would be simple, since we would have to have all the necessary information.

Therefore, at stage E, calculations of the expected likelihood value (expectation of the likelihood) are made, including latent variables, as if we were able to observe them. In step M, the maximum values of parameters' likelihood are calculated (maximum likelihood estimates) of the parameters using the maximization of the expected likelihood values obtained in step E. Next, the algorithm again performs step E using the parameters obtained in step M and so on.

A whole series of such algorithms was developed, based on the algorithm of maximizing the expectation [21-23]. For example, the structural algorithm for maximizing the mathematical expectation (structural EM algorithm) combines a standard algorithm for maximizing the mathematical expectation to optimize parameters, and an algorithm for the structural search of a selection model.

This algorithm builds networks using penalty probabilistic values, which include values, derived from Bayesian information criteria, the principle of minimum description length, and others.

Noisy-MAX node is made up of a child node *Y*, accepting n_Y values that may be tagged from 0 to $n_Y - 1$, and *N* parents, $Pa(Y) = \{X_1, \ldots, X_N\}$, representing causes of *Y*. If $X_i = 0$, it means the absence of X_i. If all the reasons are missing, the result is also missing, then the Noisy-MAX are determined according to the formula [8]:

$$P\left(Y = 0 \mid X_i = 0_{[\forall_i]}\right) = 1 \tag{1}$$

If the degrees' maximum produced by X acted independently, the degree reached by Y is determined by the formula:

$$P(Y \le y \mid x) = \prod_i P\left(Y \le y \mid \boldsymbol{X}_i = x_i, X_j = 0_{\left[\forall_{j, j \ne i}\right]}\right) \tag{2}$$

x represents a special combination of the parents of *Y*, $x = (x_1, ..., x_N)$. The probabilities that the result will take a certain value of *y*, in the case when X_i equals a certain value of x_i, provided that all the other reasons for *Y* are absent , are the parameters for the link $X_i \to Y$:

$$c_y^{x_i} = P\left(Y = y \mid X_i = x_i X_j = 0_{\left[\forall_{j, j \ne i}\right]}\right) \tag{3}$$

If X_i takes n_{X_i} values, so according to formula 1, the amount of parameters required for the link $X_i \to Y$ will be $\left(n_{X_i} - 1\right) \times \left(n_Y - 1\right)$. This model needs a reference to only one parameter, if all the variables included in Noisy are binary. We can determine the new parameters using the following formula:

$$C_y^{x_i} = P\left(Y \le y \mid X_i = x_i, X_j = 0_{\left[\forall_j, j \ne i\right]}\right) = \sum_{y'}^{y} c_y^{x_i} \ , \tag{4}$$

the formula 2 can be represented as:

$$P(Y \le y \mid x_1, \ldots, x_n) = \prod_i C_y^{x_i} \ , \tag{5}$$

CPT can be obtained given the following conditions:

$$P(y \mid x) = \begin{cases} P(Y \le 0 \mid x) & \text{if } y = 0 \\ P(Y \le y \mid x) - P(Y \le y - 1 \mid x) & \text{if } y > 0 \end{cases} \tag{6}$$

5 Experiments and Results

In developing the BN, the Genie 2.3 software environment was used. The initial nodes type is General, each node has 5 states from s1 to s5.

The following macroeconomic indicators for the period from 2012 (1st quarter) to 2017 (IV quarter) - 24 points were taken as experimental data for calculating the dependence of Ukraine's GDP growth from the volume of investment in the transport industry:

x1 - the volume of investments in land and pipeline transport of actual prices;
x2 - the volume of investment in water transport;
x3 - the volume of investment in air transport;
x4 - the volume of investments in warehousing and auxiliary activities in the field of transport;
x5 - the volume of investments in postal and courier activities.

The variety of available data can be divided into two sets: 16 measurements is training sample A, 8 measurements is test sample B (Table 1).

Table 1. Statistical data of capital investments and gross domestic product

x1	**x2**	**x3**	**x4**	**x5**	**y**
1992,1	35,7	275,3	2820,1	20,4	292894
4378,5	26,8	176	3642,6	35,4	347842
2601,8	41,7	133,5	3581,9	168	389213
3691,8	28,4	195,2	4140,5	172,4	381289
587,8	6,3	98,6	1779,2	5	303753
1189,4	37,3	137,6	2042	5,1	354814
1440,2	31,8	131,6	3039	8,2	398000
1650,3	21,4	155,3	3929,4	202,1	408631
590,3	29,3	73,3	1876,8	3,8	316905
1140,6	41	79,2	2230,4	4,1	382391
652,3	89	71,4	1820,4	9,1	440476
1072,6	48,8	96,2	3921,9	105,5	447143
1805,6	23,1	116,3	904,4	4,9	375991
874,6	96,2	193,7	1935,6	2,8	456715
2386,9	111,6	125,6	2083,2	5,3	566997
2150,9	102,9	223,1	3058,7	72,6	588841
1952,2	36,6	99,2	1439,6	13,3	455298
2256,5	38	177,7	2114,6	22,5	535701
3365,5	53,3	218,4	2646,5	18,7	671456
7383	106,2	202,3	2527,1	66,5	722912
3693	50,8	210,2	1454	6,6	591008

4181,9	55,4	260,4	1979	51,8	664760
4627,7	56,9	372,2	2852,1	53,7	833130
9455,6	74,6	340,4	5640,6	285,3	894022

At the first stage of the available in the GeNie2.3 Academic software environment, the methods of structural learning select the appropriate method. Figure 1 presents the results of selection.

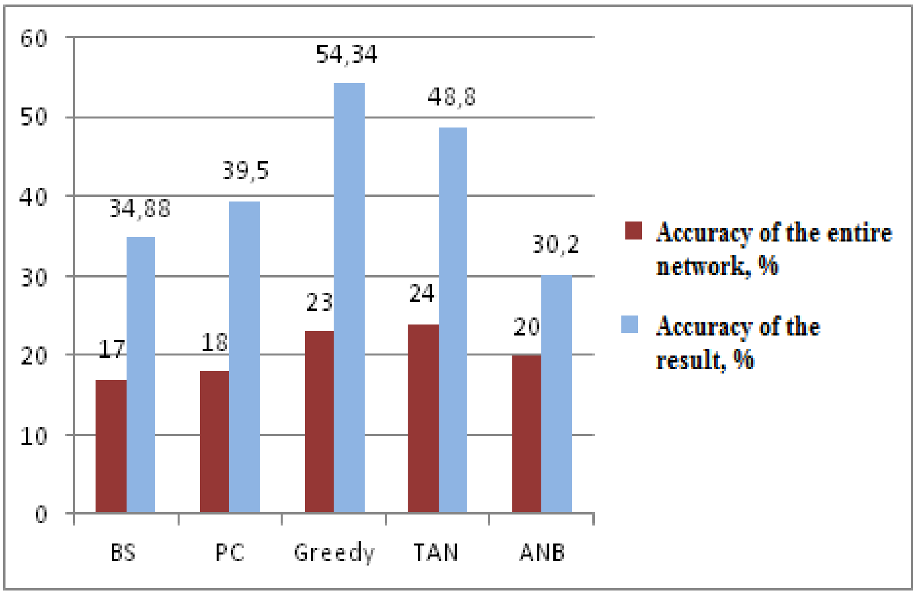

Fig. 1. Selection of a structured learning method

As a result of the experiment, a Bayesian network consisting of 6 nodes was obtained. After parametric learning, primary validation was carried out. The model has achieved 22.9% level of accuracy during the test. After revalidation, the accuracy of the model was 54.34%.

In the next step, we changed the type of all nodes to Noisy with five states from s1 to s5, the resulting node Y. The network remains the same, the data file also does not change. We conduct parametric learning, primary validation and sensitivity analysis. Comparison of accuracy is presented in the table 2:

Table 2. Comparison of accuracy after primary validation and revalidation.

	The initialaccuracy		**Accuracy after last validation**	
	Overall network accuracy,%	Accuracy of the result ,%	Overall network accuracy ,%	Accuracy of the result,%
General nodes	22,53	23,00	22,83	54,34
Noisy-max nodes	25,46	54,34	27,83	54,35

Next, we apply Hirerical discretization method. We use the Greedy algorithm, we repeat all the steps: structural training, parametric training, validation, sensitivity analysis and re-validation. the accuracy of the result remained unchanged 54,34%. Comparison of accuracy after changing the sampling method is given in table 3.

We apply Hirerical discretization method now to Noisy nodes, then using the Greedy algorithm, we repeat all the steps: structural learning, parametric learning, validation, sensitivity analysis and repeated validation. The accuracy of the entire network decreased slightly - 21.21%, but the accuracy of the result was higher by 12.32% and amounted to 66.66%.

Table 3. Comparison of accuracy after changing the discretization method.

	Discretization method ***Weigher***		Discretization method ***Hirerical***	
	Overall network accuracy,%	Accuracy of the result ,%	Overall network accuracy ,%	Accuracy of the result ,%
General nodes	22,53	54,34	22,83	54,34
Noisy-max nodes	25,46	54,35	**21,21**	**66,66**

6 Discussion

During the selection of the structural learning algorithm, it was revealed that the Greedy algorithm turned out to be an adequate method when working with the existing data set.

With Noisy-max nodes, the required resulting accuracy is achieved immediately after the initial validation. This suggests that for a network with this type of nodes there is no need for sensitivity analysis and re-validation (Table 2).

When using the Hirerical discretization method, the accuracy of the result with the General nodes remains unchanged, and with Noisy-max nodes, it increases by 12.32% (Table 3).

When using General nodes, the EM execution time during the validation process was 13 seconds. For Noisy nodes, the EM algorithm spent three times as much computing time (37 seconds). On small data sets with a small BN size, such time costs can be neglected. However, as the network increases, the time costs (and hence the computing power) will be tangible and this will have to be taken into account.

7 Conclusion

Noisy-max nodes, compared to General nodes, provide relatively high initial accuracy. General nodes require a repeated validation procedure. When using the Hirerical discretization method, the accuracy of the network result with General nodes remains unchanged, and with Noisy-max nodes, it increases (in our case, by 12.32%). However, Noisy-max nodes entail an increase in time and computational costs.

In future studies, it is planned to use the dynamic Bayesian network approach in order to trace the levels of key indicators in different time slices.

References

1. Keynes, J. M.: The General Theory of Employment. In: The Quarterly Journal of Economics (1937)
2. Zvi, M., Kane, A., Bodie, A.:Essentials of Investments Fourth Edition, Jan 1, (2001)
3. Andreassen, S., Woldbye, M., Falck B., Andersen, S.K.: MUNIN - A causal probabilistic network for interpretation of electromyographic findings. In: Proceedings of the International Joint Conference on Artificial Intelligence, Milan, Italy, 366-372. (1987)
4. Castillo, E., Gutierrez, J., Hadi, A.: Sensitivity analysis in discrete Bayesian networks. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, Vol. 27(4), 412–423. (1998)
5. Cooper, G., Herskovits, E.: A Bayesian method for the induction of probabilistic networks from data, Machine Learning 9, 309-347. (1992)
6. Cheng, J., Druzdzel, M.: AIS-BN: An Adaptive Importance Sampling Algorithm for Evidential Reasoning in Large Bayesian networks. In: Journal of Artificial Intelligence Research (JAIR), Vol. 13, 155-188. (2000)
7. Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Stat. Soc. , Vol. 39(B), 1–38. (1977).
8. Díez, F., Galán, S.: Efficient computation for the noisy MAX, Int. J. Int. Syst., Vol. 18/2, 165–177. (2004)
9. Kayaalp, M., Cooper, F.: A Bayesian network Scoring Metric That Is Based on Globally Uniform Parameter Priors, 251-258. (2002)
10. Kjærulff, U., van der Gaag, J.: Making sensitivity analysis computationally efficient. In: Proceedings of the Sixteenth Conference Uncertainty in Artificial Intelligence (UAI). San Francisco, CA: Morgan Kaufmann Publishers, 317–325. (2000)
11. Henrion, M., Kanal, L., Levitt, T., Lemmer, J.: Some practical issues in constructing belief networks. In: Uncertainty in Artificial Intelligence 3, New York: Elsevier, 161-173. (1989)
12. Díez, F.: Parameter adjustment in Bayes networks. The generalized noisy OR-gate. In: Proc. 9th Conf. Uncertainty Artif. Intell., Washington, DC, 99-105. (1993)
13. Díez, F., Mira, J., Iturralde, E., Zubillaga, S.: DIAVAL, a Bayesian expert system for echocardiography. In: Artif. Intell. Med., Vol. 10/1, 59-73. (1997)

14. Pradhan, M., Provan, G., Middleton, B., Henrion, M.: Knowledge engineering for large belief networks. In: Proc. 10th Annu. Conf. UAI, San Francisco, CA, 484-490. (1994)
15. Shwe, M., Middleton, B., Heckerman, D., Henrion, M., Horvitz, E., Lehmann, H., Cooper, G.: Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base: I. The probabilistic model and inference algorithms. In: Methods Inf. Med.,Vol. 30/ 4, 241-255. (1991)
16. Onisko, A., Druzdzel, M., Wasyluk, H.: Learning Bayesian network parameters from small data sets: Application of noisy-OR gates. In: Int. J. Approx. Reason., Vol. 27/2, 165–182. (2001)
17. Zhang, N., Poole, N.:Exploiting causal independence in Bayesian network inference. In: J. Artif. Intell. Res., Vol. 5/1, 301-328. (1996)
18. Beinlich, N. , Suermondt, N. , Chavez, N. , Cooper, G: The ALARM monitoring system: A case study with two probabilistic inference techniques for belief networks, In: Proceedings of the Second European Conference on Artificial Intelligence in Medicine, London, England, 247-256. (1989)
19. Darwiche, N.: A differential approach to inference in Bayesian networks. In Uncertainty in Artificial Intelligence: Proceedings of the Sixteenth Conference (UAI 2000), San Francisco, CA: Morgan Kaufmann Publishers, 123–132. (2000)
20. Friedman, N.: The Bayesian structural EM algorithm / Fourteenth conference on Uncertainty in Artificial Intelligence (UAI'98), Madison, Wisconsin, USA, 24–26 July, 1998. – SF.: Morgan Kaufmann, 129-138. (1998)
21. Henrion, M.: Some practical issues in constructing belief networks. In: Kanal L.N., Levitt T.S. and Lemmer J.F. (Eds.), Uncertainty in Artificial Intelligence 3 (North-Holland, Amsterdam, 161-173. (1989)
22. van der Gaag, L. C., Coup´ e, V. M.: Sensitivity analysis for threshold decision making with Bayesian belief networks. In Lamma, E., and Mello, P., eds., AI*IA 99: Advances in Artificial Intelligence, Lecture Notes in Artificial Intelligence. Berlin: Springer-Verlag. 37 – 48. (2000)
23. Zhang, Z., Kwok, J., Yeung D.: Surrogate maximization (minimization) algorithms for AdaBoost and the logistic regression model. In: Proceedings of the twenty-first international conference on machine learning (ICML), 117 p. (2004)