УДК 519.254

COMPARATIVE EFFECTIVENESS OF PARALLEL AND RECURRENT CALCULATIONS IN COMBINATORIAL ALGORITHMS OF INDUCTIVE MODELLING

Serhiy Yefimenko

International Research and Training Centre of Information Technologies and Systems of the NAS and MES of Ukraine

syefim@ukr.net

У роботі виконано порівняльне дослідження двох способів підвищення ефективності комбінаторного алгоритму МГУА – розпаралелювання обчислень за допомогою кластерних систем та рекурентного обчислення параметрів моделей. Проведено тестові експерименти з порівняння часу виконання відповідних алгоритмів, які показали високу ефективність рекурентного алгоритму.
Ключові слова: індуктивне моделювання, комбінаторний алгоритм МГУА, рекурентне обчислення параметрів моделей, паралельні обчислення, кластерна система

The paper investigates comparative effectiveness of parallel implementation and recurrent parameters estimation in combinatorial GMDH algorithm. The test experiments on run-time comparison of these two approaches for enhancing the efficiency of combinatorial algorithm are carried out. The results of these experiments confirm effectiveness of the recurrent algorithm.
Keywords: i*nductive modelling, combinatorial GMDH algorithm, recurrent parameters estimation, parallel computing, cluster system*

В работе выполнено сравнительное исследование двух способов повышения эффективности комбинаторного алгоритма МГУА – распараллеливания вычислений с помощью кластерных систем и рекуррентного вычисления параметров моделей. Проведены тестовые эксперименты по сравнению времени выполнения соответствующих алгоритмов, которые показали высокую эффективность рекуррентного алгоритма.
Ключевые слова: индуктивное моделирование, комбинаторный алгоритм МГУА, рекуррентное вычисление параметров моделей, параллельные вычисления, кластерная система

1 Introduction

The tools of mathematical statistics (such as regression and factor analysis and so on) allow solving problems of complex systems modeling. But the effectiveness of their use essentially depends on the knowledge of a priori information. Therefore it is wise to use inductive approaches allowing to speed up modeling process due to automatization of the best model search. The principal feature of GMDH (as well-known inductive method) consists in revealing (on the base of data set information) of hidden relationships, applying principle of automatic generation, successive selection of complicated models structures and external adjunction. And it is quite important that we need not set model structure as opposed to mentioned approaches.

The investigation of combinatorial GMDH algorithm follows. Appropriateness

of the combinatorial algorithm use is caused by the fact that other enumeration variants don't guarantee an optimal result which gives an exhaustive search. Computational complexity of parameters estimation stage in combinatorial algorithms is exponential by arguments amount. That is why available computer resources have to be applied completely. In this case it is advisable to use:

– the high-speed methods of parameters estimation based on the recurrent algorithms for solving of linear equations systems [1];

– paralleling of computing using multiprocessing cluster systems [2].

In the paper comparative effectiveness of parallel and recurrent calculations in combinatorial algorithms of inductive modeling with the use of computational experiment on cluster system will be investigated.

2 Combinatorial GMDH Algorithm

Let we have sample $W=(X\ y)$, dim $W=n\times(m+1)$, where X – design matrix of m input vectors, dim $X=n\times m$, y – output vector, dim $y=n\times 1$.

The problem of structural identification consist in building optimal model f^* from the set $\Im$ of models of the form $\hat{y}_f = f(X, \hat{\theta}_f)$, minimizing the value of a given criterion $CR(\cdot)$

$$f^* = \arg\min_{f\in\Im} CR(y, f(X, \hat{\theta}_f)). \quad (1)$$

We will use m-dimensional structural vector $d = \{d_1,..., d_m\}$, $d_i = \{0, 1\}$, defining set of input vectors from design matrix X to be included in the model.

The number of all possible structural vectors, could be built for m input variables, is calculated by formula

$$P_m = \sum_{s=1}^{m} C_m^s = 2^m - 1. \quad (2)$$

If the arguments number is more than 30, the exhaustive search for the acceptable time is often impossible when using even state-of-the-art personal computers.

3 Recurrent Gauss Algorithm for Parameters Estimation using Least-Squares Method

Let we have the system of n conditional equations with m unknowns:

$$X\theta = y. \quad (3)$$

The normal system $H\theta = g$ with elements

$$H = X^T X = \left\{ h_{ij},\quad i,j=\overline{1,m} \right\},\ g = X^T y = \left\{ g_i,\quad i=\overline{1,m} \right\} \quad (4)$$

corresponds to system (3).

The first step solution looks like [3]:

$$g_s^1 = \frac{g_s^0}{h_{s\,s}^1},\ \theta_1^1 = g_1^1;\quad h_{1\,i}^1 = \frac{h_{1\,i}^0}{h_{11}^0},\ \ i = \overline{2,\,m}. \tag{5}$$

Computing formulas for step s ($s = \overline{2,\,m}$) (when adding argument s to the system containing s-1 arguments) will be as follows.

Direct motion:

$$h_{i\,s}^s = (h_{i\,s}^{s-1} - \sum_{j=1}^{i-1} h_{i\,j}^{s-1} h_{j\,s}^{s-1}) / h_{i\,i}^{s-1},\quad i = \overline{2,\,s-1}, \tag{6}$$

$$h_{s\,i}^s = h_{s\,i}^{s-1} - \sum_{j=1}^{i-1} h_{s\,j}^{s-1} h_{j\,i}^{s-1},\quad i = \overline{2,\,s},\quad g_s^s = (g_s^{s-1} - \sum_{i=1}^{s-1} h_{s\,i}^{s-1} g_i^{s-1}) / h_{s\,s}^{s-1}. \tag{7}$$

Counter motion:

$$\theta_s^s = g_s^{s-1},\quad \theta_i^s = g_i^{s-1} - \sum_{j=i+1}^{s} h_{i\,j}^{s-1} \theta_j^s,\quad i = s-1,...,1. \tag{8}$$

Let's find quantitative assessment of operations number when computing of regression coefficients with recurrent Gauss algorithm. It equals $2s^2–s$ for direct motion and $s^2–s$ for counter motion when adding argument s to the system containing s-1 arguments. As is well known, dependence of computational complexity (number of elementary arithmetic operations) of regression coefficients calculation on arguments amount has cube character ($2s^3+s^2$, [4]) for classic Gauss (nonrecurrent) algorithm.

It is significant to note that theoretical assessment is rather approximative and does not allow comparing recurrent with nonrecurrent algorithms on processing speed (it will be done later). It only enables to draw conclusion about quadratic and cube dependence of computational complexity on arguments amount for recurrent and nonrecurrent algorithms, respectively.

The table 1 represents obtained quantitative assessment of computational complexity.

Tab.1

Assessment of computational complexity for regression coefficients calculation.

Gauss algorithm	*Operations amount*
Classic nonrecurrent	$2s^3+s^2$
Recurrent	$3s^2–2s$

4 Optimal Paralleling of Nonrecurrent Gauss Algorithm

We will consider the scheme of algorithm with successive complication of structures of binary numbers generator. The scheme of algorithm paralleling for determination of the initial state of binary structural vector by position for every processor of multiprocessing cluster system is presented in [5]. It provides the even loading on all multiprocessors. Paralleling of other parts for GMDH combinatorial algorithm using cluster system presents no substantial difficulties.

Let we have m arguments and κ multiprocessors within cluster. We will write down the sequence of operations for the models of complication i, $i = \overline{1, m}$:

1. Calculation of amount of combinations – $C_m^i - 1$.

2. Determination of the initial state of binary vector d for every multiprocessor j, $j = \overline{1, k}$ as a decimal number $\left[\frac{C_m^i - 1}{k}\right](j-1)+1$.

3. Conversion from the decimal number to appropriate binary number for every processor:

$$\text{position} = \left[\frac{C_m^i - 1}{k}\right](j-1)+1;$$

u=i-1, d=m-1, $C = C_d^u$;

Cycle on l, $l = \overline{1, m}$

if position<=C then $b[l]$=1, u= u -1, d= d -1, $C = C_d^u$;

else $b[i]$=0, position = position – C, u= u -1, $C = C_d^u$.

5 Results of Experiments

In test experiments we measured and compared run-time of:

– optimal parallel implementation (on the cluster system [6]) of combinatorial GMDH algorithm with nonrecurrent parameters estimation;

– combinatorial GMDH algorithm with recurrent parameters estimation on single processor.

The experiment was executed as follows: the design matrix X of size 45×25 (45 records for 25 arguments) was generated. Vector y was formed in this way: $y=x_{11}+x_{12}+x_{13}+x_{14}+x_{15}$. Time of parametric identification was measured for design matrix containing first 21 arguments, 22 arguments, …, 25 arguments.

Figure 1 represents dependence of program run-time on amount of used processors and arguments. The last row of diagram corresponds to algorithm with recurrent calculations in case of one processor using.

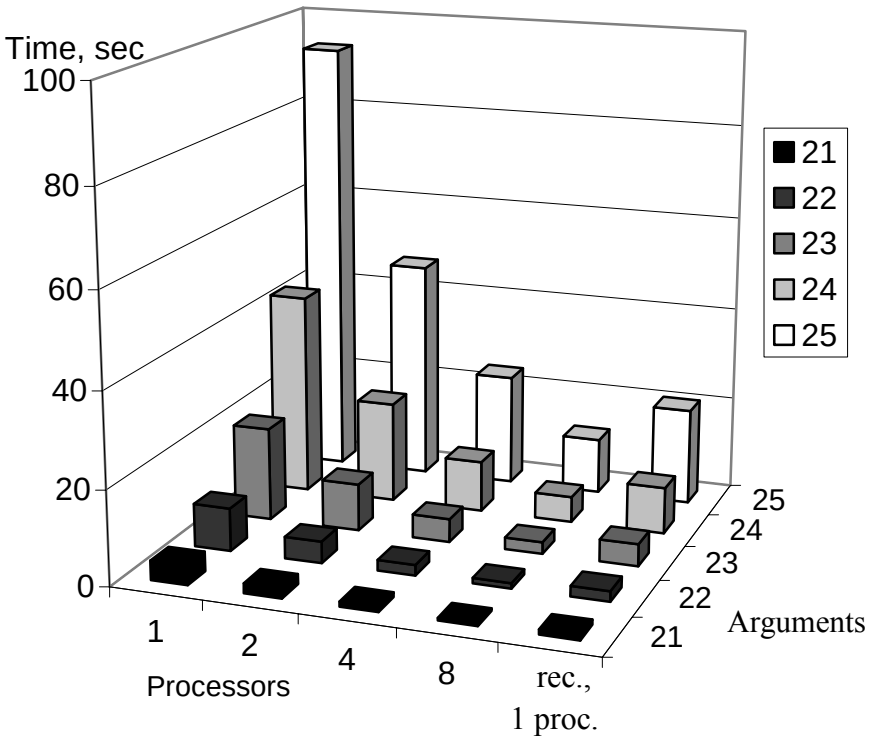

Fig. 1. Run-time of combinatorial algorithm

The effectiveness of combinatorial algorithm with recurrent parameters estimation was calculated as run-time ratio of recurrent and nonrecurrent algorithms on single processor for given arguments amount *s* (see figure 2).

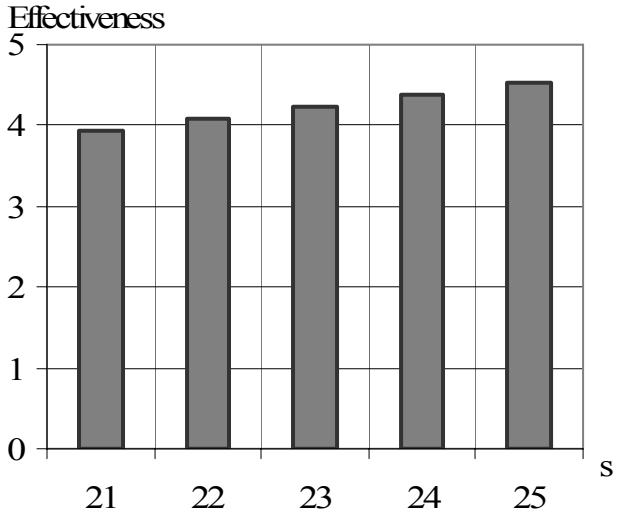

Fig. 2. Effectiveness of recurrent algorithm

6 Conclusion

The paper considers comparative effectiveness (theoretical and experimental) of parallel and recurrent calculations in combinatorial GMDH algorithm.

The results of test experiments demonstrate:

– five times superiority of recurrent calculations on nonrecurrent parameters estimation in the problem of modeling from data observed with exhaustive search of variants (for number of arguments, equal 25);

– growing effectiveness of recurrent parameters estimation with arguments amount increase.

References

1. V. S. Stepashko, and S. N. Efimenko. Sequential Estimation of the Parameters of Regression Model // Cybernetics and Systems Analysis, Springer New York, July, 2005, Vol. 41, Num. 4, pp.631-634.

2. Stepashko V.S., Yefimenko S.M., Rozenblat O.P., Chernyack A.I. On application of parallel computations in tasks of modelling on the basis of inductive approach // Problems in programming. – 2006. – № 2-3. – PP. 170-176. (in Ukrainian).

3. Yefimenko S., Stepashko V. Recurrent Gauss Algorithm for Sequential Estimation of Regression Model Parameters // Proceedings of IWIM 2011. 4nd International Workshop on Inductive Modeling, July 4-10, 2011, Kyiv. – P. 119-122.

4. N. S. Bakhvalov, N. P. Zidkov, and G. M. Kobelkov. Numerical Methods. Nauka, Moscow, 1987 (in Russian).

5. Stepashko V.S., Yefimenko S.M. Optimal paralleling for solving combinatorial modelling problems // Proceedings of the 2nd International Conference on Inductive Modelling ICIM 2008. – Kyiv, 2008. – P. 172-175.

6. http://icybcluster.org.ua