

УДК 681.5.015

МГУА КАК ИНСТРУМЕНТ АНАЛИЗА ВЗАИМОСВЯЗЕЙ МНОГОМЕРНОГО ПРОЦЕССА

Ал.В.Павлов¹, Ан.В.Павлов²

¹Национальный технический университет Украины «КПИ»,

²Международный научно-учебный центр информационных технологий и систем
НАН и МОН Украины,

Alexander_mk@ukr.net, Me_ovechka@bigmir.net

В роботі запропонований підхід для дослідження складних систем, що базується на визначенні напрямку впливу між процесами за допомогою алгоритмів самоорганізації. Запропоновані процедури фільтрації змінних на вході алгоритму моделювання, визначення множин екзогенних та ендогенних змінних системи, синтезу структури взаємозв'язків системи процесів. Запропонований підхід дозволяє врахувати можливість зміни знаку причинності у часі випередження та значно зменшити час синтезу моделей системи.

Ключеві слова: причинно-наслідковий зв'язок, напрям впливу, МГУА, система процесів, синтез структури.

New approach for complex systems analysis based on discovering of direction impact between processes using self-organizing algorithms is proposed. Input factors filtering, exogenous and endogenous factors sets determining and system processes relationships structure synthesis procedures are offered. Technique proposed makes it possible to reveal possibility of causality sign reversing between processes in the look-ahead period and significantly decrease time of the models system synthesis.

Key words: cause-effect relation, influence direction, GMDH, system of processes, structure synthesis.

В работе предложен подход к исследованию сложных систем, основанный на определении направлений влияния между процессами с помощью алгоритмов самоорганизации. Предложены процедуры фильтрации переменных на входе алгоритма моделирования, определения множества экзогенных и эндогенных переменных системы, синтеза структуры связей системы процессов. Предложенный подход позволяет учесть возможность смены знака причинности во времени упреждения и значительно снизить время синтеза системы моделей.

Ключевые слова: причинно-следственная связь, направление влияния, МГУА, система процессов, синтез структуры.

Введение. Особое место алгоритмов самоорганизации при исследовании структуры взаимосвязей сложной системы процессов обусловлено свойством моделей, получаемых в теории МГУА (метод группового учета аргументов [1]) – свойством оптимальной сложности. Благодаря этому, при исследовании системы, предметом синтеза являются не самые точные модели, а модели оптимальной сложности, что открывает новые методологические возможности для исследования взаимосвязей, решения задач анализа и синтеза структуры сложных систем.

Постановка задачи. Исследуется система из m взаимосвязанных непрерывных процессов x_1, x_2, \dots, x_m , для которой предполагается возможность её описания некоторым множеством дискретных моделей. Измерения производятся в конкретные моменты времени с наперёд заданной частотой дискретизации

Δ , данные представляются в виде матрицы X ($\dim X = n \times m$) дискретного многомерного векторного процесса:

$$X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{pmatrix},$$

задающей процессы $x_j, j = 1, \dots, m$ в отсчётах $i = 1, \dots, n$.

Необходимо исследовать систему F с целью синтеза структуры моделей прогноза.

Анализ последних исследований и публикаций. В работе [2] для решения задачи построения системы разностных уравнений оптимальной сложности предлагается генерация $M = C_m^1 + C_m^2 + \dots + C_m^m = 2^m - 1$ систем с постепенным возрастанием количества уравнений и выбор оптимальной с помощью минимизации системного критерия несмещённости решений. Для построения каждого из уравнений используется комбинаторный алгоритм МГУА. Если же перебор велик – используется комбинаторный алгоритм с «занулением» переменных или многорядные алгоритмы с пороговыми самоотборами лучших промежуточных переменных.

В работе [3], развивающей подход [2], предлагается последовательное применение критериев селекции: вначале критерия $n_{cm}^* \rightarrow \min$ для отбора F наиболее несмещённых систем уравнений, а затем критерия сходимости пошагового интегрирования $i(N)$ для выбора одной наилучшей системы уравнений. Обращается внимание на то, что рассмотренный алгоритм генерации систем уравнений не осуществляет полный перебор всех возможных систем. Действительно, имея 13 переменных и 3 запаздывания, полное уравнение имеет вид:

$$x_{i(+1)} = f(a_0, x_{1(0)}, x_{1(-1)}, x_{1(-2)}, x_{1(-3)}, x_{2(0)}, x_{2(-1)}, x_{2(-2)}, x_{2(-3)}, \dots, x_{13(0)}, x_{13(-1)}, x_{13(-2)}, x_{13(-3)}).$$

В общей сложности мы имеем $13 \cdot 2^{52}$ полинома ($i = 1, \dots, 13$), из которых необходимо составить системы с 1-м, 2-мя, ..., 13-ю уравнениями и оценить их по системному критерию несмещённости решений.

Работа [4] посвящена анализу и дальнейшему развитию алгоритма [2-3], с целью синтеза неавтономных систем уравнений и применения их в задачах управления. Отмечается, что подавляющее большинство систем, самоорганизующиеся по алгоритму в [2-3] оказываются *автономными*, т.е. не содержат независимых переменных среди своих уравнений. Такие системы не содержат *экзогенных* переменных. В работе на языке теории графов формулируются необходимые и достаточные условия для построения неавтономных моделей. Разработан алгоритм синтеза неавтономных систем. Рассчитана верхняя оценка объёма перебора алгоритма и предлагается её сокращение при помощи использования некоторого аналога градиентных методов поиска минимума системного критерия.

$$x_{k+s,j} = f_j(x_{k,i}, x_{k-1,i}, \dots, x_{k-l,i}), \quad (4)$$

где Δ – шаг дискретизации непрерывного процесса; s – глубина прогноза (интервал упреждения); l – количество запаздываний (лагов). Все дальнейшие процедуры рассматриваются при заданных (или полученных путем направленного перебора при моделировании) фиксированных значениях параметров Δ , s , l .

Для формализации процедуры сравнения двух конкурирующих моделей (3) и (4) определим следующие характеристики:

Пусть δ_i – погрешность моделирования x_i по x_j , δ_j – погрешность моделирования x_j по x_i .

Определим *индикативную погрешность* δ_i^I , указывающую индексом i направления влияния связи между процессами, следующим образом:

$$\min\{\delta_i, \delta_j\} = \delta_i = \delta_i^I \Leftrightarrow x_j \rightarrow x_i, \quad (5)$$

$$\min\{\delta_i, \delta_j\} = \delta_j = \delta_j^I \Leftrightarrow x_j \leftarrow x_i. \quad (6)$$

Отметим, что эконометрический причинно-следственный анализ применяет схожую с (5), (6) процедуру, однако используя при этом только линейные аналоги моделей (3), (4) принципиально сужает возможности подхода.

Также близкая по идее процедура определения направления влияния была предложена в [7]. При этом сравнивались конкурирующие модели вида:

$$x_{k+1,i} = f_i(x_{k,i}, x_{k-1,i}, \dots, x_{k-l,i}, x_{k,j}, x_{k-1,j}, \dots, x_{k-l,j}), \quad (7)$$

$$x_{k+1,j} = f_j(x_{k,j}, x_{k-1,j}, \dots, x_{k-l,j}, x_{k,i}, x_{k-1,i}, \dots, x_{k-l,i}),$$

используя гармонический критерий экзогенности. Из анализа работы [7] очевидно, что гармонический критерий сравнивает следствия взаимодействия только самых низкочастотных по времени составляющих исследуемого процесса, кроме того, алгоритм чрезвычайно нагружен емкими, по требуемым вычислительным ресурсам, процедурами, что делает его применение крайне затруднительным в реальных задачах. Отметим, что при построении модели (7), в отличие от (3), пользуясь правилами (5), (6), мы будем оценивать искаженное (а не «чистое») влияние процессов x_j , x_i друг на друга, поскольку в модели (7) в правой части присутствуют переменные собственных запаздываний. Как показал численный эксперимент, такое искажение, как правило, приводит к снижению показателя надежности принятия решения о направлении связи. О показателе скажем ниже.

Порог индикативной погрешности δ^* определим как предельно большое значение δ^I , при котором допускается возможность принятия решения о направлении односторонней связи между переменными x_i и x_j .

Показатель надёжности принятия решения о направлении влияния между парой переменных x_i и x_j определим:

$$\Omega_{ij} = |\delta_i - \delta_j|.$$

Порог показателя надёжности Ω^* определим как предельно малое значение Ω_{ij} , при котором допускается возможность принятия решения о направлении односторонней связи между переменными x_i и x_j .

Для возможности сравнения результатов в разных парах переменных в одной задаче системного синтеза полезно ввести относительный показатель надежности.

Относительным показателем надежности η_{ji} принятия решения о направлении связи между переменными $x_j \rightarrow x_i$ назовем отношение:

$$\eta_{ji} = \frac{\Omega_{ij}}{\delta_i^I}.$$

Минимальное допустимое значение относительного показателя надежности (порог относительного показателя) определим как:

$$\eta^* = \frac{\Omega^*}{\delta^*}.$$

При условии использования нормированной ошибки δ , величину относительного показателя надежности η_{ij} легко переместить в единый диапазон. Например, для приведения её в $[0,100]$, следует воспользоваться формулами:

$$\eta_{ji} = \frac{\Omega_{ij}}{0,01 + \delta_i^I}, \quad \eta^* = \frac{\Omega^*}{0,01 + \delta^*}.$$

Выбор порогов δ^* , Ω^* индивидуален для каждой задачи. Позже, при рассмотрении процедуры фильтрации переменных на входе алгоритма моделирования, мы рассмотрим, к чему может привести не оптимальный выбор значений параметров δ^* , Ω^* .

Отметим, что указанный выше признак (5), (6) определения направления влияния между исследуемыми процессами является лишь необходимым, но не достаточным условием для открытия *физических* причинно-следственных связей между процессами. Действительно, условия (5) и (6) являются не достаточными по следующим причинам:

- а) имеет место простое совпадение поведения переменных;
- б) переменные x_i и x_j являются следствиями одной общей причины (третьей переменной), и как следствия, проявляются с различными запаздываниями. Однако, когда целью моделирования есть прогноз, то ни причина а), ни причина б), не являются препятствием для включения в модель фиктивной связи, при условии, что учет данной переменной дает улучшение прогноза системы. Поэтому в данной работе мы говорим о причинно-следственных связях в статистическом смысле этого слова.

Для иллюстрации действенности данной методики рассмотрим пример на реальных данных. Сравним модели (3) и (4), полученные алгоритмом МГУА МАКСО [8] для пары переменных: x_1 – совокупные денежные доходы населения в Украине (млн. грн.) и x_2 – совокупные денежные расходы и сбережения населения в Украине (млн. грн.). Данные усреднены на интервале месяц за период с ноября 1995 по сентябрь 1999 года. Указанные переменные выбраны в виду очевидности для нас направления действующей причинно-следственной

связи в данной паре. Параметры моделирования при использовании алгоритма МАКСО были следующими: $s = 1, l = 10, N_A = 27, N_B = 22, N = 49, \Delta = 1$ месяц.

На рис.1 приведены графики процессов x_1 и x_2 . Графики моделей прогноза вида (3) и (4), полученные при моделировании x_1 по x_2 приведены на рис.2, x_2 по x_1 – на рис.3.

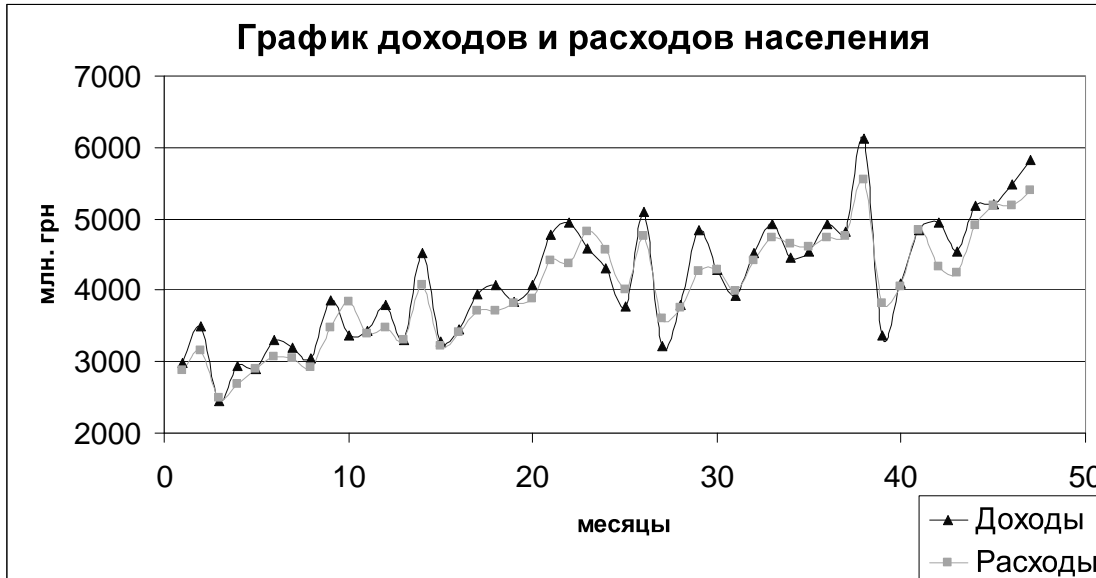


Рисунок 1. Графики доходов и расходов населения.

По графикам рисунка 1 очевидно, что кривая расходов (x_2) в целом бежит впереди кривой доходов (x_1), то есть характерные по времени повторения расходов происходят позже (правее на графике), чем доходов, что согласуется со здравым смыслом. Тот же вывод следует из анализа точности моделей x_1 и x_2 вида (3), (4), графики которых приближают табличные данные – рис.2 и рис.3.

Модель оптимальной сложности, полученная для прогноза x_1 по переменной x_2 :

$$\hat{x}_{1,k+1} = 6339,09 - 8349360 \cdot x_{2,k-5} \cdot x_{2,k}^{1/3} \cdot x_{2,k-8}^{-1} \cdot x_{2,k-2}^{-1} \cdot x_{2,k-9}^{-1/3} + 87,3463 \cdot x_{2,k-7} \cdot x_{2,k-1} \cdot x_{2,k-3}^{1/3} \cdot x_{2,k-10}^{-1} \cdot x_{2,k-4}^{-1} - 14,5702 \cdot x_{2,k-6} \cdot x_{2,k-1} \cdot x_{2,k-10}^{-2/3} \cdot x_{2,k-2}^{-1/3} - 5262,91 \cdot x_{2,k}^{2/3} \cdot x_{2,k-2}^{1/3} \cdot x_{2,k-8}^{-2/3} \cdot x_{2,k-9}^{-4/3} - 5130,33 \cdot x_{2,k-9} \cdot x_{2,k-7}^{1/3} \cdot x_{2,k-9}^{1/3} \cdot x_{2,k-2}^{-1} \cdot x_{2,k-3}^{-1}.$$

Нормированная среднеквадратичная ошибка (NMSE) на рабочих точках:

$$\delta_1 = NMSE_{раб} = \sqrt{\frac{\sum_{i=1}^N (x_{1i} - \hat{x}_{1i})^2}{\sum_{i=1}^N (x_{1i} - \bar{x}_1)^2}} = 0,73$$

Модель оптимальной сложности, полученная для прогноза x_2 по переменной x_1 :

$$\hat{x}_{2,k+1} = 8313,01 - 18056100 \cdot x_{1,k-1}^{1/3} \cdot x_{1,k-2}^{1/3} \cdot x_{1,k-5}^{1/3} \cdot x_{1,k-7}^{1/3} \cdot x_{1,k-8}^{1/3} \cdot x_{1,k-4} \cdot x_{1,k-7} \cdot x_{1,k-10} \cdot x_{1,k}^{-1} \cdot x_{1,k-3}^{-1} - 21384,29 \cdot x_{1,k-2}^{1/3} \cdot x_{1,k-9}^{1/3} \cdot x_{1,k-6}^{-1/3} \cdot x_{1,k-1}^{-1/3} \cdot x_{1,k}^{-1/3} - 0,0275 \cdot x_{1,k-2} \cdot x_{1,k-3} \cdot x_{1,k-10}^{1/3} \cdot x_{1,k}^{-1} \cdot x_{1,k-4}^{-1/3} - 0,01916 \cdot x_{1,k}^3 \cdot x_{1,k-1}^{-1} \cdot x_{1,k-10}^{-1}.$$

$$\delta_2 = NMSE_{раб} = 0,53298.$$

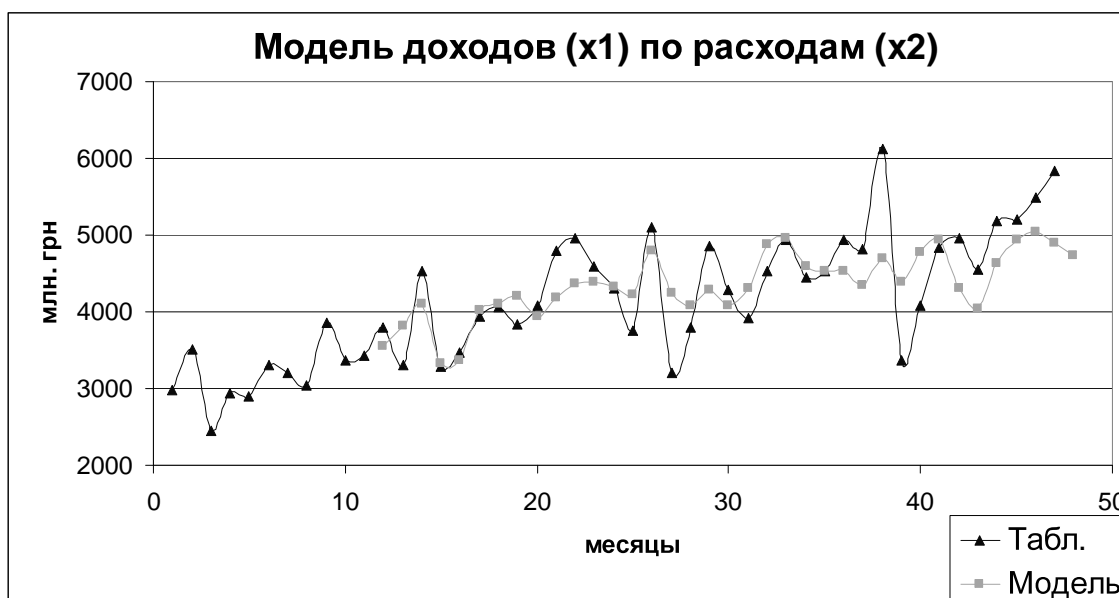


Рисунок 2. Моделирование доходов по расходам.



Рисунок 3. Моделирование расходов по доходам.

Отметим, что более точная модель x_2 с $\delta_2 = 0,53$ приближает x_2 по x_1 , поэтому определяем причиной – x_1 , следствием – x_2 , то есть, в данной паре переменных расходы являются следствием доходов: $x_1 \rightarrow x_2$. Пороговые значения

приняты на уровне $\delta^* = 0,9$, $\Omega^* = 0,1$, то есть разность точности в конкурирующих моделях для принятия решения о направлении связи должна превышать 10%, притом, что ошибка лучшей из конкурирующих моделей должна быть не более 0,9. Значение пороговой ошибки δ^* выбрано достаточно большой, поскольку в конкурирующих моделях (3) и (4) оценивается лишь небольшая часть реально действующего многофакторного причинно-следственного влияния. Результат был получен при $\delta_2^I = 0,53$ и $\Omega_{12} = 0,19$, при этом $\delta_2^I < \delta^*$ и $\Omega_{12} > \Omega^*$. Как видим, в данном примере процедура принятия решения о направлении связи с большим запасом удовлетворяет выбранным ограничивающим параметрам.

Доопределим остальные связи между переменными, учитывая введенные характеристики:

$$\begin{array}{ll} 1) x_j \rightarrow x_i, \Leftrightarrow \delta_i < \delta_j, \delta_i \leq \delta^*, \Omega_{ij} \geq \Omega^* & 2) x_j \leftarrow x_i, \Leftrightarrow \delta_i > \delta_j, \delta_j \leq \delta^*, \Omega_{ij} \geq \Omega^* \\ 3) x_j \leftrightarrow x_i, \Leftrightarrow \delta_i < \delta_j, \delta_i \leq \delta^*, \Omega_{ij} < \Omega^* & 4) x_j \leftrightarrow x_i, \Leftrightarrow \delta_i < \delta_j, \delta_i > \delta^*, \Omega_{ij} < \Omega^* \\ & \delta_i > \delta_j, \delta_j \leq \delta^*, \Omega_{ij} < \Omega^* & \delta_i > \delta_j, \delta_j > \delta^*, \Omega_{ij} < \Omega^* \end{array}$$

Знак \leftrightarrow означает, что связь между процессами x_i , x_j отсутствует.

В указанных 4-х случаях будем говорить о надёжном принятии решения о направлении связи между процессами x_i и x_j .

Последний вариант – невыясненность отношения между переменными:

$$5) x_j ? x_i, \Leftrightarrow \delta_i < \delta_j, \delta_i > \delta^*, \Omega_{ij} \geq \Omega^* \\ \delta_i > \delta_j, \delta_j > \delta^*, \Omega_{ij} \geq \Omega^*$$

Данный вариант предполагает дальнейшее исследование.

Рассмотренная процедура сравнения конкурирующих моделей может иметь ряд полезных приложений. Приведём некоторые из них:

- определение множества эндогенных и экзогенных переменных системы;
- фильтрация переменных по направлению связи;
- построение ориентированного графа системы.

Определение множества эндогенных и экзогенных переменных системы. Сформулируем алгоритм определения множества $X^{\text{экз}}$ экзогенных переменных системы: для каждой переменной x_i строим все пары (x_i, x_j) конкурирующих моделей (3), (4), $x_j \in X \setminus \{x_i\}$, и если переменная x_i во всех парах имеет направление $x_i \rightarrow x_j$ или $x_i \leftrightarrow x_j$, то она добавляется в множество $X^{\text{экз}}$. Тогда для описания объекта нужно моделировать множество эндогенных (внутренних) переменных: $X^{\text{энд}} = X \setminus X^{\text{экз}}$.

Процедура фильтрации переменных по направлению связи. Предложим процедуру первичной (до моделирования) фильтрации переменных для моделирования каждой x_i , $x_i \in X^{\text{энд}}$. Данная процедура формирует множество переменных X^{x_i} , являющихся претендентами на вход при моделировании x_i . В соответствии с 5-ю выше определёнными вариантами исхода в процедуре сравнения пары конкурирующих моделей (x_i, x_j) , $x_i, x_j \in X^{\text{энд}}$ и $x_j \in X \setminus \{x_i\}$ обозначим следующие множества:

$$\{x_j\}^+ \text{ – множество переменных, имеющих направление } x_j \rightarrow x_i;$$

$\{x_i\}^-$ – множество переменных, имеющих направление $x_j \leftarrow x_i$;
 $\{x_i\}^0$ – множество не связанных с x_i переменных $x_j \leftrightarrow x_i$;
 $\{x_i\}^\pm$ – множество переменных, имеющих двунаправленную связь $x_j \leftrightarrow x_i$;
 $\{x_i\}^?$ – множество переменных, для которых не решен вопрос о направлении связи.

Проанализируем, каким образом неоптимальный выбор порогов влияет на процедуру сравнения конкурирующих моделей (3) и (4).

Пусть $x_j \rightarrow x_i$, $x_j \in \{x_i\}^+$, тогда уменьшая δ^* ↓, имеем $x_j ? x_i$, $x_j \in \{x_i\}^?$, далее $\Omega^* \uparrow$, имеем $x_j \leftrightarrow x_i$, $x_j \in \{x_i\}^0$, увеличивая δ^* ↑, имеем $x_j \leftrightarrow x_i$, $x_j \in \{x_i\}^\pm$, и снова $\Omega^* \downarrow$, имеем $x_j \rightarrow x_i$, $x_j \in \{x_i\}^+$. Как очевидно, увеличение требований по надёжности (δ^* ↓, $\Omega^* \uparrow$) приведёт к вымыванию переменных из множеств $\{x_i\}^+$ и $\{x_i\}^-$, и присоединению их к множествам $\{x_i\}^0$, $\{x_i\}^\pm$, $\{x_i\}^?$. Снижение требований надёжности приведёт к исключению переменных из множеств $\{x_i\}^0$, $\{x_i\}^\pm$, $\{x_i\}^?$ и включению их в множества $\{x_i\}^+$, $\{x_i\}^-$. Таким образом, неоптимальный выбор порогов не приводит к ошибкам при определении направления связи, а лишь к увеличению или уменьшению учтенных связей с x_i . Следовательно, предлагается два варианта фильтрации переменных:

1) «жесткая» фильтрация: на вход x_i поступает множество переменных $X^{x_i} = \{x_i\}^+ \cup \{x_i\}^\pm$, отфильтровывается множество $\{x_i\}^- \cup \{x_i\}^0 \cup \{x_i\}^?$.

2) «мягкая» фильтрация: на вход x_i поступает множество переменных $X^{x_i} = \{x_i\}^+ \cup \{x_i\}^\pm \cup \{x_i\}^0 \cup \{x_i\}^?$, отфильтровывается множество $\{x_i\}^-$.

Анализ последствий неоптимального выбора порогов процедуры сравнения моделей позволяет сделать вывод о том, что завышенные требования по надёжности к процедуре фильтрации могут быть компенсированы «мягкостью» варианта её процедуры.

Аналогичным образом можно определить «мягкое» и «жесткое» определение множеств экзогенных и эндогенных переменных системы.

Построение ориентированного графа системы. Рассмотренный алгоритм определения направления действующих связей, дает информацию для задания графа структуры системы матрицей смежности. Результаты анализа представляются матрицей Q (рис. 4):

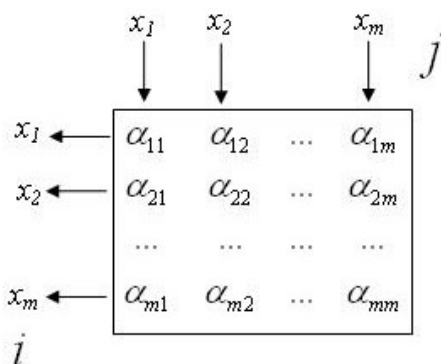


Рисунок 4. Матрица смежности.

Определим элементы матрицы Q в зависимости от варианта процедуры фильтрации.

«Жесткая» фильтрация:

$$\begin{cases} \alpha_{ji} = 1, \text{ если } x_j \in \{x_i\}^+ \cup \{x_i\}^\pm, i \neq j \\ \alpha_{ji} = 0, \text{ если } x_j \in \{x_i\}^- \cup \{x_i\}^0 \cup \{x_i\}^?, i \neq j \end{cases}$$

«Мягкая» фильтрация:

$$\begin{cases} \alpha_{ji} = 1, \text{ если } x_j \in \{x_i\}^+ \cup \{x_i\}^\pm \cup \{x_i\}^0 \cup \{x_i\}^?, i \neq j \\ \alpha_{ji} = 0, \text{ если } x_j \in \{x_i\}^-, i \neq j \end{cases}$$

Определенный таким образом ориентированный граф, отвечает структуре связей исследуемой системы процессов и дает информацию для определения входов и выходов при её моделировании с целью получения наилучшего прогноза.

Приведём пример применения изложенной методики для построения ориентированного графа некоторой подсистемы $\{x_1, x_2, x_3, x_4, x_5\}$ экономических процессов в Украине, где x_1 – денежные доходы населения (млн. грн.), x_2 – денежные расходы и сбережения (млн. грн.), x_3 – индексы цен производства легкой промышленности (%), x_4 – объем производства промышленности всего (млн. грн.), x_5 – объем производства легкой промышленности (млн. грн.).

При анализе и моделировании использовались значения переменных за период с ноября 1995г. по май 1999г. Будем последовательно перебирать пары переменных x_i, x_j , строить для них конкурирующие модели вида (3), (4) и, по лучшей точности модели в паре, принимать решение о направлении причинно-следственной связи между ними. Применяя версию МАКСО [8] комбинаторно-селекционного алгоритма МГУА синтезируем модели оптимальной сложности вида (3), (4) при этом $s = 1, l = 10$ месяцев, $\Delta = 1$ месяц, $\delta^* = 0,9; \Omega^* = 0,1$. Конкретный вид моделей и графиков опустим, а будем приводить только соответствующие значения $NMSE$.

Включаем в рассмотрение переменную x_2 :

$x_1 \rightarrow x_2, NMSE = 0,73; x_2 \rightarrow x_1, NMSE = 0,53$, вывод: $x_2 \rightarrow x_1$.

Включаем в рассмотрение переменную x_3 :

$x_2 \rightarrow x_3, NMSE = 0,7; x_3 \rightarrow x_2, NMSE = 0,89$, вывод: $x_2 \rightarrow x_3$;

$x_1 \rightarrow x_3, NMSE = 0,76; x_3 \rightarrow x_1, NMSE = 0,88$, вывод: $x_1 \rightarrow x_3$.

Включаем в рассмотрение переменную x_5 :

$x_5 \rightarrow x_1, NMSE = 0,7; x_1 \rightarrow x_5, NMSE = 0,93$, вывод: $x_5 \rightarrow x_1$;

$x_5 \rightarrow x_2, NMSE = 0,71; x_2 \rightarrow x_5, NMSE = 0,93$, вывод: $x_5 \rightarrow x_2$;

$x_5 \rightarrow x_3, NMSE = 0,43; x_3 \rightarrow x_5, NMSE = 0,64$, вывод: $x_5 \rightarrow x_3$.

Включаем в рассмотрение переменную x_4 :

$x_1 \rightarrow x_4, NMSE = 0,47; x_4 \rightarrow x_1, NMSE = 0,65$, вывод: $x_1 \rightarrow x_4$;

$x_2 \rightarrow x_4, NMSE = 0,47; x_4 \rightarrow x_2, NMSE = 0,68$, вывод: $x_2 \rightarrow x_4$;

$x_3 \rightarrow x_4, NMSE = 0,6; x_4 \rightarrow x_3, NMSE = 0,72$, вывод: $x_3 \rightarrow x_4$;

$x_5 \rightarrow x_4, NMSE = 0,84; x_4 \rightarrow x_5, NMSE = 0,74$, вывод: $x_4 \rightarrow x_5$.

Приведенные результаты позволяют сформировать матрицу смежности Q (рис.5) для задания ориентированного графа (рис.6):

	x_1	x_2	x_3	x_5	x_4
x_1 ←	×	0	0	1	0
x_2 ←	1	×	0	1	0
x_3 ←	1	1	×	1	0
x_5 ←	0	0	0	×	1
x_4 ←	1	1	1	0	×

Рисунок 5. Матрица смежности подсистемы.

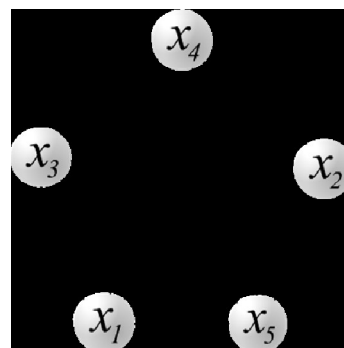


Рисунок 6. Ориентированный граф подсистемы.

Поскольку при значениях порогов, заданных на уровне $\delta^* = 0,9$; $\Omega^* = 0,1$ множества $\{x_i\}^- \cup \{x_i\}^0 \cup \{x_i\}^?$, $i = 1, \dots, m$ оказались пустыми, то нет разницы, какой из вариантов фильтрации использовать, свидетельствуя о том, что пороги могли быть выбраны с заниженной надёжностью. На завершающем этапе синтеза структуры объекта необходимо промоделировать вершины графа, которые определяют собственно систему уравнений прогноза рассматриваемых процессов:

$$x_{k+1,1} = f_1(x_{k,1}, \dots, x_{k-l,1}, x_{k,5}, \dots, x_{k-l,5});$$

$$x_{k+1,2} = f_2(x_{k,1}, \dots, x_{k-l,1}, x_{k,2}, \dots, x_{k-l,2}, x_{k,5}, \dots, x_{k-l,5});$$

$$x_{k+1,3} = f_3(x_{k,1}, \dots, x_{k-l,1}, x_{k,2}, \dots, x_{k-l,2}, x_{k,3}, \dots, x_{k-l,3}, x_{k,5}, \dots, x_{k-l,5});$$

$$x_{k+1,4} = f_4(x_{k,1}, \dots, x_{k-l,1}, x_{k,2}, \dots, x_{k-l,2}, x_{k,3}, \dots, x_{k-l,3}, x_{k,4}, \dots, x_{k-l,4});$$

$$x_{k+1,5} = f_5(x_{k,4}, \dots, x_{k-l,4}, x_{k,5}, \dots, x_{k-l,5}).$$

Данная процедура вносит окончательную коррекцию в структуру связей и модель системы процессов.

Анализ выше сформированной структуры может вызывать замечания. Например, полученное направление связей между переменной x_4 – объем производства промышленности и переменными, x_1 – денежные доходы населения и x_2 – денежные расходы и сбережения населения: $x_1 \rightarrow x_4$, $x_2 \rightarrow x_4$ на первый взгляд противоречит здравому смыслу. Однако, при увеличении учтенных запаздываний до 20-ти месяцев для моделей переменных x_1 , x_4 получаем $NMSE_4 = 0,53$ и $NMSE_1 = 0,335 \Rightarrow x_4 \rightarrow x_1$, то есть знак направления зависимости инвертируется относительно полученного на коротких (до 10-ти месяцев) лагах запаздываний. Следовательно, на модели с короткими лагами мы получаем известный эффект влияния инфляционных процессов на экономические показатели, а на длинных лагах получаем противоположное направление зависимости: доходы населения зависят от объема производства. Для тех же условий для моделей x_2 , x_4 имеем $NMSE_2 = 0,31$ и $NMSE_4 = 0,31$. Таким образом, получаем аргументы для дополнительного исследования системы на предмет определения двунаправленности связи для переменных x_2 , x_4 на лагах до 20-ти месяцев.

Фильтрация списка переменных для моделирования индекса потребительских цен в Украине. При анализе и моделировании использованы значения переменных x_1, x_2, \dots, x_{22} за период с ноября 1995г. по май 1999г. Где x_1 – денежные доходы населения (млн. грн.); x_2 – денежные расходы и сбережения (млн. грн.); x_3 – индекс потребительских цен (%); x_4 – индексы цен в промышленности (%); x_5 – индексы цен в легкой промышленности (%); x_6 – индексы ВВП (%); x_7 – объем производства промышленности всего (млн. грн.); x_8 – розничный товароборот всего (млн. грн.); x_9 – общая занятость (тыс. чел.); x_{10} – занятость в легкой промышленности (тыс. чел.); x_{11} – заработная плата (ном., грн.); x_{12} – индекс реальной зарплаты ("нетто" %); x_{13} – зарплата в легкой промышленности (ном.,грн.); x_{14} – установлена средняя ставка рефинансирования комбанков.(% годовых); x_{15} – средняя процентная ставка банковской системы на кредиты в национальной валюте (% годов.); x_{16} – денежная масса наличности в обращении (млн.грн.); x_{17} – курс гривни к \$100 США (грн.); x_{18} – дебиторская задолженность между предприятиями. (млн.грн.); x_{19} – кредиторская задолженность между предприятиями (млн.грн.); x_{20} – расходы Сводного бюджета, всего (млн.грн.); x_{21} – доходы Сводного бюджета, всего (млн.грн.); x_{22} – объем производства легкой промышленности.(млн.грн.);

Фильтрация переменных осуществлялась при следующих параметрах алгоритма МАКСО: $l = 10$, $N_A = 27$, $N_B = 19$, $N_C = 3$. При этом $\Delta = 1$ месяц, $\delta^* = 0,9$; $\Omega^* = 0,1$.

Результаты фильтрации. Для моделирования x_3 – индекса потребительских цен были отобраны следующие аргументы со следующими значениями порогов:

x_2	$\delta_3 = 0,88$; $\delta_2 = 0,96$	x_8	$\delta_3 = 0,87$; $\delta_8 = 1,03$
x_{13}	$\delta_3 = 0,89$; $\delta_{13} = 1,06$	x_{14}	$\delta_3 = 0,9$; $\delta_{14} = 0,1$
x_{18}	$\delta_3 = 0,78$; $\delta_{18} = 1,04$	x_{19}	$\delta_3 = 0,82$; $\delta_{19} = 1,03$
x_{20}	$\delta_3 = 0,83$; $\delta_{20} = 0,99$	x_{21}	$\delta_3 = 0,8$; $\delta_{21} = 0,96$

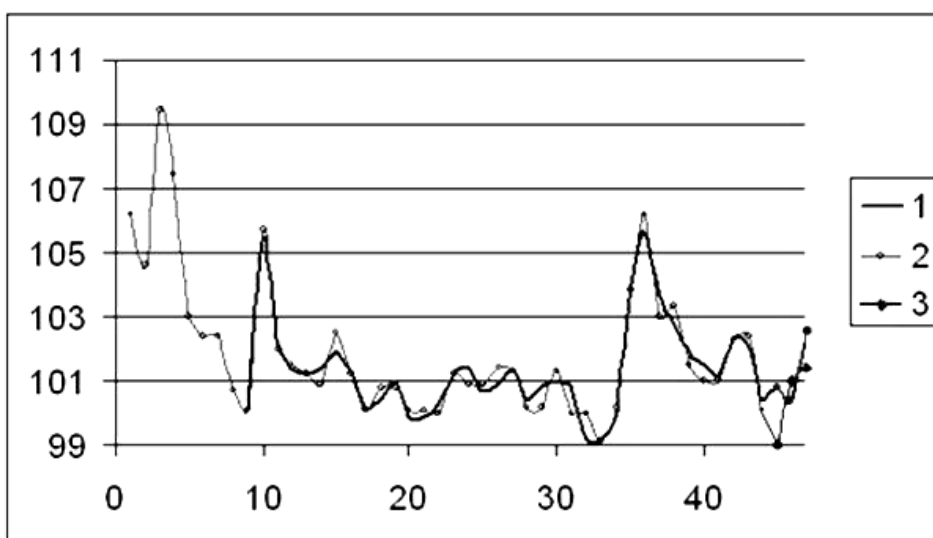


Рисунок 7. График модели индекса потребительских цен.
1 – график модели, 2 – график табличных точек, 3 – точки экзамена

Вид модели прогноза для переменной x_3 на три шага:

$$\begin{aligned} \hat{x}_{3,k+3} = & 99,175 - 0,000023 \cdot x_{22,k} \cdot x_{21,k-5}^2 \cdot x_{8,k} + 1003909769 \cdot x_{14,k} \cdot x_{2,k}^{-1} \cdot x_{21,k-4}^{-2} + \\ & + 13991031 \cdot x_{13,k-1} \cdot x_{21,k-5}^{-2} \cdot x_{22,k-1}^{-1} + 0,034 \cdot x_{14,k-5} \cdot x_{20,k-5} \cdot x_{14,k}^{-1} \cdot x_{22,k-2}^{-1} - 87212948 \cdot \\ & \cdot x_{8,k-2} \cdot x_{19,k-4}^{-1} \cdot x_{20,k-5}^{-1} \cdot x_{21,k}^{-1} + 27754585087 \cdot x_{13,k-3}^{-1} \cdot x_{20,k-5}^{-1} \cdot x_{20,k-4}^{-1} \cdot x_{13,k-1}^{-1}. \end{aligned}$$

Средний модуль ошибки RE на экзамене (нормированный к диапазону изменения переменной в выборке экзаменационных данных)

$$RE = \frac{\sum_{i=1}^{N_C} |\hat{x}_i - x_i|}{N_C \cdot |x_{\max} - x_{\min}|}$$

Полученные оценки для многофакторной модели переменной x_3 равны соответственно: $\delta_3 = NMSE_{\text{раб}} = 0,24902$, $RE = 0,1175$. Точность модели практически не отличается от характеристик модели, полученной при использовании всех 22 аргументов, при этом количество аргументов поданных на вход алгоритма моделирования снижено на 63%, время работы алгоритма с учетом процедур определения направления связей сокращено в 3 раза.

В заключение отметим, что предложенный подход к структурному синтезу уравнений системы прогноза, принципиально отличается от подходов, рассмотренных в [2-7]. В данной работе оптимальная система синтезируется на основании анализа направления связей между переменными, в отличие от анализа степени плотности связи, как это делается в [2-7].

Предложим обобщение изложенной методики исследования системы процессов для различных значений времени упреждения s .

Исследование «причинно-следственных» связей во времени упреждения. Приводя соотношения (3), (4) мы говорили о, найденных перебором общих для обеих моделей структур оптимальной сложности, параметрах Δ , l и s . И если, значение параметра Δ (минимальный шаг дискретности) чаще всего мы наследуем из готовой выборкой данных, параметр l (количество запаздываний) может быть выбран из условия учета предполагаемой минимальной частоты с ненулевой амплитудой в данных (частота тренда), то с параметром s (упреждением) связано, в общем случае, определенное усложнение анализа причинности рассматриваемых процессов.

Рассмотрим возможность изменения направления влияния (изменение значения одной ячейки матрицы Q) между любыми двумя процессами x_i , x_j во времени упреждения s . В примере построения графа связей системы мы однозначно определили направление связи в каждой ячейке при $s = 1$, ожидая, что данное направление сохранится и при больших значениях s (рис. 8). Толщина коридора, наложенного на нижнюю кривую, соответствует величине выбранного порога Ω^* .

Однако в общем случае направление связи может меняться во времени упреждения, т.е. процессы взаимно влияют друг на друга, меняя тем самым знак причинно-следственных отношении на противоположный (рис. 9). Данный вариант соответствует более сложным взаимодействиям переменных x_i, x_j , амплитуды которых, по-разному суммируются на различных частотах и поэтому по-разному влияют на значения моделей и ошибок при различных упреждениях s .

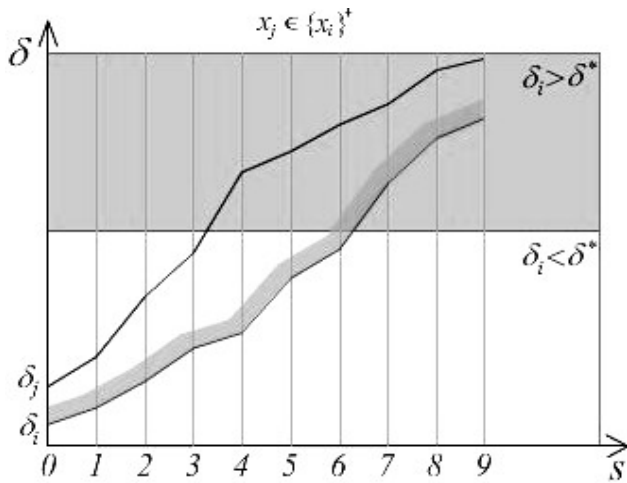


Рисунок 8. Однонаправленная связь между процессами во времени упреждения

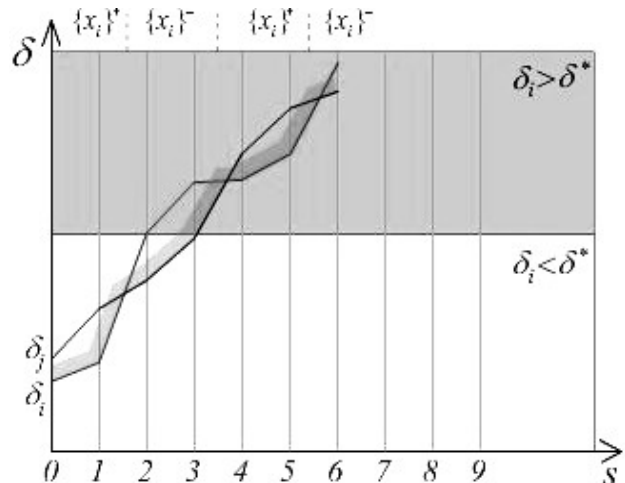


Рисунок 9. Общий случай взаимодействия процессов во времени упреждения

Исследуем более подробно множество $\{x_i\}^\pm$. Как видно из рисунка 10, $x_j \in \{x_i\}^\pm$ может быть частным случаем перехода знака направления связи на противоположный ($s = 2, s = 6$). Поэтому множество $\{x_i\}^\pm$ должно быть обязательно доисследовано во времени упреждения и переменная x_j отфильтрована при возникновении указанного частного случая. В этом случае множество $\{x_i\}^\pm$ будет содержать только те переменные, для которых условие $\Omega_{ij} < \Omega^*, \delta_i < \delta^*$ сохраняется для нескольких значений s подряд (рис. 11).

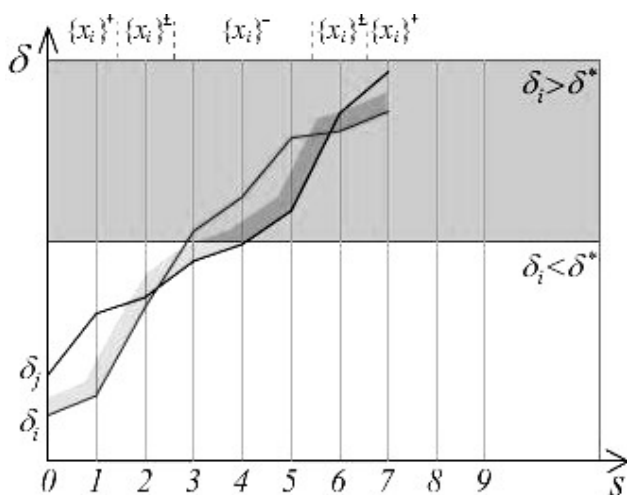


Рисунок 10. Двухнаправленная связь как частный случай перехода знака влияния на противоположный

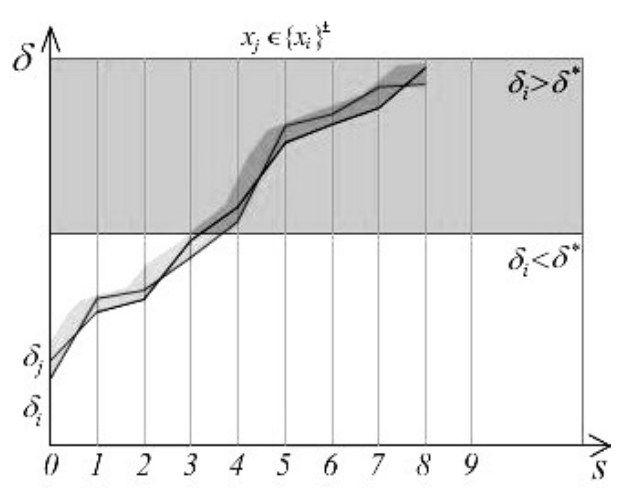


Рисунок 11. Двухнаправленная связь между процессами во времени упреждения

Для каждого s , $s = 1, \dots$ может быть получена своя матрица смежности Q_s , представляющая механизм взаимодействия переменных на соответствующем интервале упреждения.

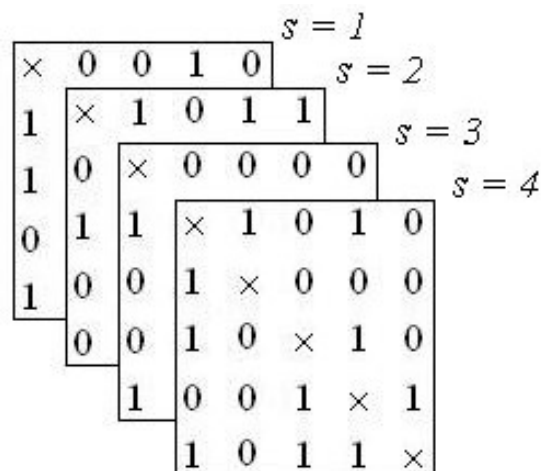


Рисунок 12. Наложение матриц Q_s .

Для интегральной картины взаимодействия воспользуемся наложением матриц смежности для каждого s , $s = 1, \dots$ (рис. 12), получив суммарную матрицу Q , по которой можно объективно судить как о наличии экзогенных переменных в системе для всех s , так и о суммарных взаимодействующих влияниях.

В заключение отметим возможные направления доработки предложенной методики:

1. Сформулируем рекомендации по выбору характерных параметров δ^* и Ω^* . Требования надёжности принятия решения требуют $\delta^* \downarrow$, $\Omega^* \uparrow$. С другой стороны, наличие значительного числа переменных во множестве $\{x_i\}^\pm$ свидетельствует о возможности снижения требований по надёжности ($\Omega^* \downarrow$, $\delta^* \uparrow$). Также снижение требований по надёжности целесообразно осуществлять пропорционально величине предполагаемой многофакторности полных зависимостей, в которых участвуют переменные. Наличие значительного уровня шумов в данных требует относительного увеличения порога $\delta^* \uparrow$.

Как видим, рекомендации противоречивы, поэтому в дальнейшем необходимо предложить некоторую процедуру по оптимальному подбору параметров δ^* и Ω^* .

2. Отметим возможность переноса, при необходимости, принципов изложенного исследования на статистические причинно-следственные зависимости второго и более порядков: $x_i = f(x_j, x_k)$, $x_i = f(x_j, x_k, x_p)$, ..., и так далее, вплоть до полной зависимости, что очевидно, уже не имеет рационального смысла.

Выводы.

Предложен подход к исследованию сложных систем, основанный на определении направлений влияния между процессами с помощью алгоритмов самоорганизации. Его применение позволяет осуществить фильтрацию переменных на входе алгоритма моделирования, определить множество экзогенных и эндо-

генных переменных системы, синтезировать структуру связей системы процессов. Предложенный подход позволяет учесть возможность смены знака причинности во времени упреждения и значительно снизить время синтеза системы моделей.

Список литературы

[1] Ивахненко А.Г., Степашко В.С. Помехоустойчивость моделирования.— Киев: «Наук.думка», 1985, - 216 с

[2] Стоун Р., Ивахненко О.Г., Висоцький В.М., Сьоміна Л.П. Відкриття законів складних об'єктів у випадку, коли вихідні змінні не вказані. – Автоматика, 1977, №6, с. 27–40.

[3] Стоун Р., Ивахненко А.Г., Высоцкий В.Н., Сёмина Л.П. Открытие законов сложных объектов в случае, когда выходные переменные не указаны. Ч.2. Структурная идентификация. – Автоматика, 1979, №1, с. 25–37.

[4] Ивахненко А.Г., Белозерский Е.А., Козубовский С.Ф. Об открытии законов сложных систем в случае, когда выходные переменные не указаны. Часть III. Самоорганизация неавтономных систем уравнений. – Автоматика, 1979, №4, с. 15–23.

[5] Ивахненко А.Г., Костенко Ю.В., Системный анализ и долгосрочный прогноз квазистатических систем на основе самоорганизации моделей. Часть I. Системный анализ на уровне трендов. – Автоматика, 1982, №3, с. 11–19.

[6] Ивахненко А.Г., Костенко Ю.В., Голеусов И.В. Системный анализ и долгосрочный прогноз квазистатических систем на основе самоорганизации моделей. Часть II. Объективный анализ без априорного указания внешних воздействий. – Автоматика, 1983, №3, с. 3–11.

[7] Ивахненко А.Г., Копа Ю.В., Костенко Ю.В. Системный анализ и долгосрочный прогноз квазистатических систем на основе самоорганизации моделей. Часть III. Разделение выходных переменных по степени экзогенности для восстановления закономерности объекта моделирования. – Автоматика, 1984, №4, с.10–17.

[8] Ванін В.В., Павлов О.В. Розробка та застосування алгоритмів самоорганізації для моделювання складних процесів та об'єктів, що відображаються точковим каркасом / Праці Таврійської державної агротехнічної академії. Вип.4, Том 24, Мелітополь, 2004. – с. 51–56.