УДК 630.383

DOI: <http://dx.doi.org/10.20914/2310-1202-2016-1-45-51>

Доцент В.Г. Козлов,

(Воронеж. гос. ун-т. инж. технол.) кафедра информационной безопасности.
тел. 89103434840

доцент И.Н. Журавлев,

(Воронеж. гос. лесотехн. университет им. Г.Ф. Морозова) кафедра безопасности жизнедеятельности. тел. 89103434840

профессор Е.В. Кондрашова,

(Воронеж. гос. аграрн. университет им. императ. Петра I) кафедра технического сервиса и технологии машиностроения. тел. 89103434840

соискатель М.М. Умаров

(Москов. гос. университет леса) кафедра физической культуры и спорта
тел. 89103434840

E-mail: skrypnikovvsafe@mail.ru

Associate professor V.G. Kozlov,

(Voronezh state university of engineering technology), Department of information security.
phone. 89103434840

associate professor V.N. Zhuravlev,

(Voronezh state forestry engineering university named after G.F. Morozov)
Department of life safety. phone. 89103434840

professor E.V. Kondrashova,

(Voronezh state agrarian university named after Emperor Peter I) Department of technical service and engineering technology. phone. 89103434840

applicant M.M. Umarov

(Moscow state forest university) Department of physical education and sport.
phone. 89103434840

E-mail: skrypnikovvsafe@mail.ru

Математическая модель статистической идентификации информационного обеспечения автомобильного транспорта

Mathematical model of statistical identification of information support of road transport

Реферат. В данной работе на основании метода статистической идентификации, использующего теорию самоорганизующихся систем, построена многофакторная модель взаимосвязи автомобильного транспорта и его системы обучения. Исходная информация для модели представлена рядом среднегодовых параметров функционирования автомобильного транспорта и его информационного обеспечения, в том числе, комплексом параметров системы обучения (входные параметры), системы управления автомобильным транспортом и выходными параметрами. Задаются два критерия: критерий устойчивости модели и критерий корреляции. Программа определяет их минимум и находит единственную модель оптимальной сложности. В заданном количестве параметров устанавливается математическая связь каждого выходного параметра со всеми остальными. Для повышения точности и регулярности прогноза часть узлов интерполяции выделяется в проверочную последовательность данных. Остальные данные образуют обучающую последовательность. Решение модели основано на принципе селекции. Выполняется оно при постепенном усложнении математического описания и полном переборе всех возможных вариантов моделей по указанным критериям. Достоинства предложенной модели: достаточно адекватно отображает реальный процесс, позволяет вводить любые дополнительные входные параметры и определить их влияние на отдельные выходные параметры автомобильного транспорта, позволяет поочередно изменять величины основных параметров в определенной пропорции и определять соответствующие изменения выходных параметров системы автомобильного транспорта, позволяет прогнозировать выходные параметры функционирования автомобильного транспорта.

© Козлов В.Г., Журавлев И.Н.,
Кондрашова Е.В., Умаров М.М., 2016

Для цитирования

Козлов В.Г., Журавлев И.Н., Кондрашова Е.В., Умаров М.М. Математическая модель статистической идентификации информационного обеспечения автомобильного транспорта // Вестник Воронежского государственного университета инженерных технологий. 2016. №1. С. 45-51. doi:10.20914/2310-1202-2016-1-45-51.

For cite

Kozlov V.G., Zhuravlev V.N., Kondrashova E.V., Umarov M.M. Mathematical model of statistical identification of information support of road transport *Vestnik voronezhskogo gosudarstvennogo universiteta inzhenernyh tekhnologiy* [Proceedings of the Voronezh state university of engineering technologies]. 2016, no. 1, pp. 45-51. (In Russ.). doi: 10.20914/ 2310-1202-2016-1-45-51.

Summary. In this paper based on the statistical identification method using the theory of self-organizing systems, built multifactor model the relationship of road transport and training system. Background information for the model represented by a number of parameters of average annual road transport operations and information provision, including training complex system parameters (inputs), road management and output parameters. Ask two criteria: stability criterion model and test correlation. The program determines their minimum, and is the only model of optimal complexity. The predetermined number of parameters established mathematical relationship of each output parameter with the others. To improve the accuracy and regularity of the forecast of the interpolation nodes allocated in the test data sequence. Other data form the training sequence. Decision model based on the principle of selection. Running it with the gradual complication of the mathematical description and exhaustive search of all possible variants of the models on the specified criteria. Advantages of the proposed model: adequately reflects the actual process, allows you to enter any additional input parameters and determine their impact on the individual output parameters of the road transport, allows in turn change the values of key parameters in a certain ratio and to determine the appropriate changes the output parameters of the road transport, allows to predict the output parameters road transport operations.

Ключевые слова: моделирование, автомобильный транспорт, информационное обеспечение, селекция, корреляционный анализ.

Keywords modeling, road transport, information technology, selection, correlation analysis.

Введение. Сейчас в нашей стране и за рубежом, динамическим методам прогноза и, в частности, динамическому прогнозу уделяется самое пристальное внимание.

Цель работы. В данной работе на основании метода статистической идентификации, использующей теорию самоорганизующихся систем, построена многофакторная модель взаимосвязи автомобильного транспорта и его системы обучения.

Исходная информация для модели представлена рядом среднегодовых параметров функционирования автомобильного транспорта и его информационного обеспечения, в том числе, комплексом параметров системы обучения (входные параметры), системы управления автомобильным транспортом и выходными параметрами.

Задаются два критерия: критерий устойчивости модели и критерий корреляции. Программа определяет их минимум, и, таким образом, находит единственную модель оптимальной сложности. В заданном количестве параметров устанавливается математическая связь каждого выходного параметра со всеми остальными.

Для повышения точности и регулярности прогноза часть узлов интерполяции выделяется в проверочную последовательность данных. Остальные данные образуют обучающую последовательность.

Методика. Решение модели основано на принципе селекции. Выполняется оно при постепенном усложнении математического описания и полном переборе всех возможных вариантов моделей по указанным критериям. Полное описание автомобильного транспорта $\varphi = f(x_1, x_2, \dots, x_n)$ заменяется несколькими рядами частных описаний.

Первый ряд:

$$Y_1 = f(x_1, x_2); Y_2 = f(x_1, x_3); \dots; Y_S = f(x_{n-1}, x_n), \quad (1)$$

где $S = C_n^2$.

Второй ряд:

$$Z_1 = f(y_1, y_2); Z_2 = f(y_1, y_3); \dots; Z_p = f(y_{s-1}, y_n), \quad (2)$$

где $p = C_S^2$ и т.д.

Степень полного описания удваивается с каждым рядом. Каждое частное описание является функцией только двух параметров, поэтому коэффициенты частных описаний могут быть легко определены по данным обучающей последовательности.

Исключая промежуточные переменные, можно получить аналог полного описания. Последовательными отборами в каждом пороговом ряду пропускается некоторое количество только самых регулярных или несмещенных переменных [1, 2].

Степень регулярности оценивается по величине среднеквадратической ошибки, средней для всех выбираемых в каждом ряду (поколении) переменных или для одной самой точной переменной на проверочной последовательности. Критерий регулярности производительности автомобильного транспорта с аргументом φ_i рассчитывается по формуле:

$$\Delta^2(1) = \frac{\sum_{i=1}^{N_{np}} (\varphi_i - \varphi_i^*)^2}{\sum_{i=1}^{N_{np}} \varphi_i^2} 100\%, \quad (3)$$

где $\Delta(1)$ - среднеквадратическая ошибка на проверочной последовательности, %; φ_i - значение прогноза в i -й точке, $i=1,2,3,\dots, N_{np}$;

φ_i^* - действительные значения в той же точке.

Далее составляются частные описания для всех возможных пар переменных первого ряда, характеризующих функционирование автомобильного транспорта:

$$y_i = a_{0i} + a_{1j}x_j + a_{2i}x_h + a_{3i}x_jx_h + a_{4i}x_j^2 + a_{5i}x_h^2, \quad (4)$$

где $j = 1, 2, \dots, n-1$; $h = 2, 3, \dots, n$;

$i = 1, 2, \dots, C_n^2$; n - число переменных.

Количество таких уравнений будет $C_n^2 = \frac{18 \cdot 17}{2} = 153$. Из них наиболее регулярными оказались 18 уравнений.

На каждом последующем ряду также выбираются 18 самых регулярных уравнений регрессии.

Особенность прогноза, полученного при помощи метода статистической идентификации, состоит в том, что из модели могут быть исключены, на первый взгляд, очень важные переменные. Это говорит о том, что исключенные переменные могут быть косвенным образом выражены через другие переменные. Объясняется это тесной взаимосвязью всех переменных [3].

Степень устойчивости (несмещенности) оценивается по критерию устойчивости. Для его расчета все параметры ранжируются по величине дисперсии:

$$D^2 = \frac{1}{N} \sum_{i=1}^N \left[\frac{(\varphi_i - \bar{\varphi}_i)}{\bar{\varphi}_i} \right]^2. \quad (5)$$

Параметры с четными номерами образуют первую последовательность R_1 , с нечетными - R_2 . После каждого ряда селекции и определяется по F уравнений вида:

$$\begin{aligned} Y &= f(x_i x_j); \quad Z = f(y_i y_j); \quad V = f(z_i z_j); \\ W &= f(v_i v_j) \text{ и т.д.} \end{aligned} \quad (6)$$

Синтез уравнений в каждом ряду выполняется по два раза: первая последовательность является обучающей, вторая – проверочной ($R_1 = N_{об}$; $R_2 = N_{пр}$). Уравнения регрессии обозначаются $Y'_X = f(x_i x_j)$, где $1 \leq k \leq F$, затем функции последовательностей меняются местами $R_1 = N_{пр}$, $R_2 = N_{об}$, а уравнения регрессии обозначаются $Y''_X = f(x_i x_j)$. Каждое из уравнений регрессии оценивается по величине среднеквадратического отклонения, рассчитанного по всем точкам обеих последовательностей:

$$n_{уст1} = \frac{1}{R_1 + R_2} \sum_{i=1}^{R_1+R_2} (\tilde{Y}'_X - Y''_X)^2. \quad (7)$$

Выбирается F уравнений с меньшей оценкой $n_{уст1}$. Критерий устойчивости решений на первом ряду определяется как среднее значение показателя устойчивости для F отобранных устойчивых уравнений:

$$N_{уст1} = \frac{1}{F} \sum_{i=1}^F n_{устi}. \quad (8)$$

Аналогично определяются критерии устойчивости для остальных рядов селекции:

$$N_{устO} = \frac{1}{F} \sum_{i=1}^F n_{устO}, \quad (9)$$

где O - количество рядов.

Схема селекции
 $18 \rightarrow \frac{18 \cdot 17}{2} \rightarrow 18 \frac{18 \cdot 17}{2} \rightarrow \dots$ Правило остановки селекции $N_{см} \rightarrow \min$.

Таким образом, построение эконометрической модели статистической идентификации системы управления автомобильным транспортом состоит из нескольких последовательных этапов (рисунок 1).

Первый этап – сбор данных. Собираются данные функционирования всех подсистем. Отбираются показатели, отражающие основные функции автомобильного транспорта. Исходная информация представлена среднегодовыми параметрами этих подсистем, начиная с 2000 г.

Сложность управления многоуровневыми иерархическими структурами приводит к необходимости введения показателей, которые могут быть использованы на всех уровнях системы управления и изменения (укрупнения и детализации) некоторых показателей для облегчения учета, анализа и планирования работы автотранспортного предприятия, управления [4].

Вторым этапом построения эконометрической модели является качественный, логический анализ причинно-следственных связей между выбранными показателями. Устанавливаются наиболее тесные связи, строится блок-схема со всеми логическими связями между элементами и параметрами, а затем матрица связей. Матрица разбивается на блоки и подматрицы.

Третий этап – корреляционный анализ. Модель строится методом наименьших квадратов по данным за 10 лет. После выделения блоков со связями находим исходную матрицу и ищем тесноту связи между параметрами (по максимуму коэффициента корреляции) в зависимости:

$$\begin{aligned} y(T) &= x_1(T-1); \quad y(T) = x_2(T-2); \\ y(T) &= x_3(T-3), \end{aligned} \quad (10)$$

то есть находится запаздывание воздействия информационного обеспечения на параметры функционирования автомобильного транспорта.

Рисунок 1. Схема этапов построения эконометрической модели

Процесс функционирования информационного обеспечения таков, что его воздействие на автомобильный транспорт, эффективность в данный момент определяется как сумма значений этих показателей за период с начала функционирования. Началом функционирования следует считать момент поступления информации к автомобилизации, конец – моральное старение информации [4].

В первый год лишь некоторая часть информации дает эффект, причем она действует не только в данный год, но и в последующие годы. Одни виды информации сразу вступают в действие с наибольшей эффективностью, другие начинают давать максимальную отдачу через несколько лет, то есть с определенным запаздыванием, которое принято называть лагом. Нулевой лаг воздействия считается, если максимальный эффект достигается в год поступления информации, годовой лаг – через год.

Лаги воздействия на производительность труда некоторых параметров представлены на рисунке 2. Здесь x_5 говорит о тесной связи линейно зависимых величин; x_4 означает ярко выраженную зависимость производительности труда при лаге, равном 3; x_{21} обладает нулевым

лагом, связь между величинами носит марковский характер, то есть связь оперативная; зависимость x_{20} , г свидетельствует о том, что существуют два периода изменения связей двух величин: 3 и 5 лет.

Четвертый этап – отбор наиболее представительных факторов. Из рассмотренной матрицы отбираются тесно связанные параметры, наиболее активно воздействующие на аргумент $y(T)$. Причем, аргументы классифицируются по срокам запаздывания. Критерием отбора является коэффициент корреляции, равный не менее 0,5.

Программа отбирает переменные, в наибольшей степени влияющие на величину производительности труда на автомобильном транспорте. Из 79 параметров программой отобрано 18, причем остальные не исключаются из рассмотрения. Их влияние косвенным путем учитывается в факторных нагрузках основных переменных [4].

Пятый этап – факторный анализ переменных. Все элементы матрицы корреляций отличны по модулю от крайних значений, то есть от 0 и 1. Вариации отдельных признаков могут рассматриваться как результат воздей-

ствия нескольких показателей, описывающих специфику каждого признака.

Если между признаками существует корреляция, то объясняться это может двояко: либо один признак определяет все остальные, либо существует общий фактор, не включенный в рассмотрение, который оказывает значительное влияние на коррелированные признаки. Метод факторного анализа нацелен на выявление этих общих факторов.

Рисунок 2. Динамика коэффициентов корреляции с учетом лагов

Учесть влияние всех параметров одновременно практически невозможно, поэтому использован метод группового учета аргументов, согласно которому все параметры разделены на три группы. В них выделены параметры, нагрузки которых близки к единице. Данные с шумовым эффектом и переменные, вызвавшие этот шум, исключены. В первой группе выделены пять ярко выраженных параметров. Параметры с близкими нагрузками объединены и по ним построена таблица запаздывания [3].

В 39-мерном пространстве строим векторы, соответствующие факторным нагрузкам, вызываемым 39 параметрами функционирования автомобильного транспорта и его информационного обеспечения. Через них проводим плоскость так, чтобы среднеквадратическое отклонение каждого вектора от нее было минимальным и на нее проектируем все векторы.

Вращая ось координат, совмещаем ось с проекцией одного из факторов на плоскость. Таким образом, количество измерений пространства уменьшилось на единицу. Построив плоскость через 38 векторов, спроектировав их на плоскость и повернув оси координат в плоскости до совмещения оси с одной из проекций вектора, получим 37 факторов и т.д. Описанный метод главных компонент позволяет выбрать из некоторого конечного множества наиболее важные факторы. Критерием выбора является минимум разброса векторов, спроектировав их на плоскость и повернув оси координат в плоскости до совмещения оси с одной из проекций вектора, получим 36 факторов и т.д. Описанный метод главных компонент позволяет выбрать из некоторого конечного множества наиболее важные факторы. Критерием выбора является минимум разброса [5].

Отбрасывая таким образом несущественные факторы, в работе получены пять главных факторов. Как видно из полученных факторных нагрузок, тематическая направленность и объем дают большую проекцию на фактор 3 – фактор, учитывающий качество и полноту информации. Коэффициент выпуска на линию и количество автомобилей дают значительную проекцию на фактор 2 – фактор, учитывающий оснащение и качество технической эксплуатации транспортных средств. Повышение квалификации (обучение) водителей дает большую проекцию вектора на фактор 4 – фактор обучения. И, наконец, тираж и количество килобайт информации дают значительную проекцию на фактор 5. Таким образом, выделены пять факторов, определяющих производительность труда: техническую и коммерческую эксплуатацию, полноту информационного обеспечения, обучение и материальные ресурсы [1, 2].

Шестой этап – построение системы дифференциально-разностных уравнений, описывающих влияние информационного обеспечения на автомобильный транспорт. В соответствии с логической цепочкой, начиная с нижней ветви, выстраиваются несколько параметров и соединяются попарно связями. На проверочной последовательности выбирается следующий уровень. Параллельно с этим выбирается уровень временного прогнозирования по программе.

Построение модели взаимодействия автомобильного транспорта и его информационного обеспечения предполагает установление зависимости между рассматриваемыми параметрами и, как следствие, возможность определения некоторых рассматриваемых величин.

Функционирование автомобильного транспорта и его информационного обеспечения определяется n параметрами $x_1, x_2, x_3, \dots, x_n$, в каждом из которых содержится m показателей временного ряда за m лет, то есть имеем исходную матрицу $n \times m$:

$$\begin{pmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1m} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{nm} \end{pmatrix} \quad (11)$$

При построении модели предполагаем, что каждый показатель x_i в момент времени $j+1$ зависит от своего предыдущего значения x_{ij} и от всех значений других показателей x_k с учетом лага L_{xi} . Тогда значения в год $T+1$ можно выразить:

$$X_i(T+1) = f \left[\begin{matrix} X_i(T), X_1(T-L_{1i}), \dots, \\ X_{i-1}(T-L_{i-1,i}), \dots, \\ X_n(T-L_{ni}) \end{matrix} \right], \quad (12)$$

$i = 1, 2, 3, \dots, n$.

Используя методы корреляционного анализа, определяем лаги, установив максимальное $R_{\max}$ и минимальное $R_{\min}$ значения коэффициентов корреляции. Если коэффициент корреляции $r_{ij} < R_{\min}$, то случайные величины считаем некоррелированными и исключаем из рассмотрения, если $r_{ij} > R_{\max}$, то они связаны линейной зависимостью и исключаем одну из этих величин. Находим матрицу лагов. Для этого представим первую зависимость приведенного выше уравнения в виде:

$$X_i(T+1) = f \left[\begin{matrix} X_1(T), X_2(T-L_{2i}), \\ X_3(T-L_{3i}), \dots, \\ X_n(T-L_{ni}) \end{matrix} \right], \quad (13)$$

$i = 1, 2, 3, \dots, n$.

При помощи стандартной программы определяем коэффициенты корреляции между значениями X_1 и его значений, сдвинутыми на год, то есть

x_{11}	x_{12}	x_{13}	$\dots$	$x_{1,m-1}$	$x_{1,m}$	
	x_{11}	x_{12}	$\dots$	$x_{1,m-1}$	$x_{1,m}$	$x_{1,m}$

Затем находим коэффициенты корреляции r_{12} при всех возможных значениях лага L_{21} от 0 до $L_{\max}$. В расчет принимаются по указанным выше причинам значения $R_{\min} < r_{11} < R_{\max}$. За значение L_{21} выбираем то, при котором коэффициент корреляции максимальный.

Для построения функции f_i необходимо, чтобы она зависела не менее чем от двух аргументов.

Здесь возможны два предельных случая:

1. У переменной $X_i(T+1)$ нет ни одного аргумента. Тогда значения переменной $X_i(T+1)$, $X_i(T+2)$, ... могут быть получены эвристическими методами и вводятся в память ЭВМ извне [5].

2. Переменная $X_i(T+1)$ зависит только от своего предыдущего значения $X_i(T)$. Тогда прогноз проводится путем экспоненциального сглаживания:

$$X_i(T+1) = KX_i(T-1) + (1-K)X_i(T), \quad (14)$$

где K – коэффициент сглаживания, $K=0,1$.

Для нахождения коэффициентов $a_1 - a_6$, обеспечивающих наименьшее квадратическое отклонение значений y , необходимо не менее семи независимых измерений, отсюда условие:

$$m_1 \geq m - L_{\max} - 1 \geq 7. \quad (15)$$

Обозначив первую функцию через Y_1 , а ее значения в узлах интерполяции (при попарных значениях $x_{11}, x_{21}, x_{12}, x_{22}$ и т.д.) – через $y_{11}, y_{12}, \dots, y_{1m_1}$, строим зависимость от других пар аргументов (x_1, x_3) , (x_2, x_3) и т.д. Их число равно $C_{n_1}^2$. Значения сводим в матрицу. В верхней строке помещаем экспериментальные значения величин y

$$n_2 \left\{ \begin{matrix} y_1 & y_2 & y_3 & \dots & y_{m_1} \\ y_{11} & y_{12} & y_{13} & \dots & y_{1m_1} \\ y_{21} & y_{22} & y_{23} & \dots & y_{2m_1} \\ \dots & \dots & \dots & \dots & \dots \\ y_{n_2,1} & y_{n_2,2} & y_{n_2,3} & \dots & y_{n_2,m_1} \end{matrix} \right\} \quad (16)$$

После этого находим коэффициенты корреляции между элементами первой и последующих строк и ранжируем матрицу в порядке убывания коэффициентов корреляции. Затем отбрасываем $(n_2 - n_1)$ нижних строк ранжированной матрицы, получаем новую матрицу размером $n_1 \times m_1$: