

Авдеева В.М., Крючкова И.Н., Кравец О.Я.
ВЕРИФИКАЦИЯ ИНСТРУМЕНТАЛЬНЫХ СРЕДСТВ РЕАЛИЗАЦИИ
ПРОГНОЗИРОВАНИЯ НАЛОГОВЫХ ПОСТУПЛЕНИЙ

МРИ МНС РФ №1 по Воронежской области, г.Воронеж
Международный институт компьютерных технологий, г.Воронеж
Воронежский государственный технический университет

При решении задачи прогнозирования временных рядов, в качестве которых выступают объемы налоговых платежей по территории, возникает проблема корректного определения обучающего, тестирующего и контрольного множеств. При формировании указанных выборок предполагается, что значения независимых переменных и зависящей от них исследуемой величины определены в одной и той же временной точке. Однако известно множество примеров [1, 2, 3], когда влияние одного события на другое проявляется по прошествии некоторого времени, величину которого называют временным лагом Δ . В случае наличия неизвестного временного лага задача нейросетевого моделирования временных рядов существенно усложняется, так как появляется неопределенность выбора обучающих примеров.

Обычно целью нейросетевого прогнозирования является предсказание значения некоторой переменной на основе ее предыдущих значений. Обучающий пример формируется путем наложения на массив исходных данных временного «окна». Следующий пример формируется сдвигом «окна» на один временной интервал вперед.

В [4, 5] показано, что учет значений не только прогнозируемой величины, но и факторов, значимо влияющих на нее, позволяет повысить точность прогнозирования. В этом случае каждый обучающий пример содержит историю значений по факторам, значимо влияющим на значение прогнозируемой величины и историю значений самой прогнозируемой величины, а также требуемый выход нейросети [6]. Ширина временного «окна», то есть количество предыдущих значений, на основе которых формируется прогноз, может быть определено с помощью анализа графиков автокорреляционных функций входных факторов.

С учетом отсроченного влияния одного или нескольких значимых факторов на прогнозируемую величину, появляется необходимость более точно определять ширину временного окна по каждому фактору. При ширине окна меньшей, чем временной лаг, «сдвинутое» значение фактора, соответствующее значению прогнозируемой величины в определенной временной точке, не войдет в обучающий пример для этой точки.

Пусть $x(t_1), x(t_2), \dots, x(t_n)$ – значения независимой переменной, $y(t_1), y(t_2), \dots, y(t_n)$ – временной ряд, описывающий прогнозируемую величину. Предположим, что существует аналитическое выражение зависимости $Y(x)$, причем

$$Y(t) = f(x(t - \Delta)), \quad (1)$$

где Δ - временной лаг.

Для примера выбрана следующая функция

$$Y(t) = 10 \log x^2(t^*), \quad (2)$$

где $t^* = t - 3$.

Рис. 1 иллюстрирует взаимное расположение графиков $X(t)$ и $Y(t)$, рис. 2 – выбор ширины временного окна для X по графику автокорреляционной функции [4].

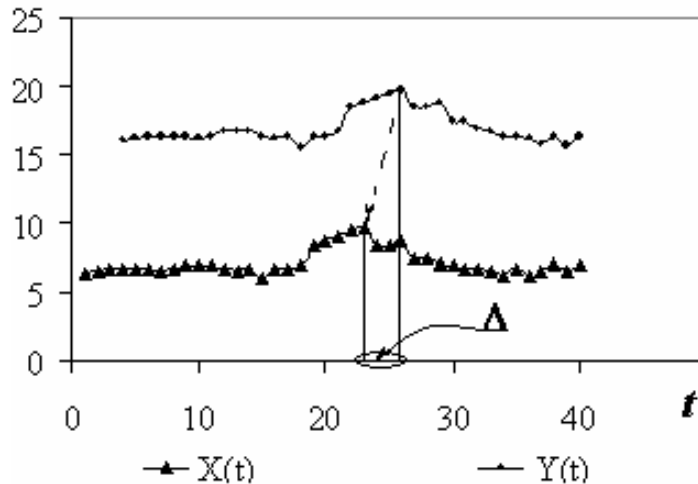


Рис. 1. Временные ряды значений независимой (X) и зависимой (Y) переменных

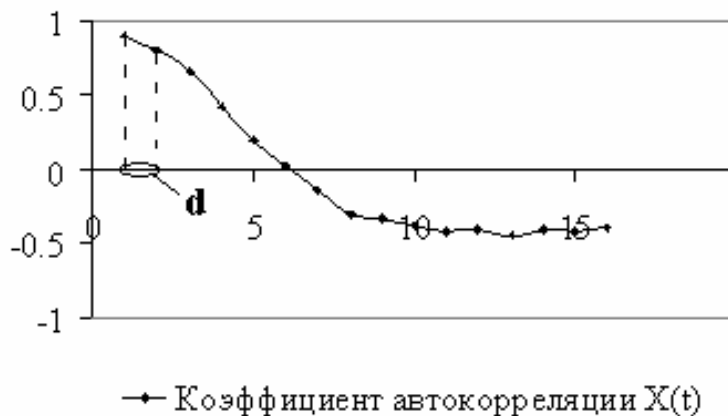


Рис. 2. График автокорреляционной функции независимой переменной X

Предположим, что вид функции $Y(t)$ неизвестен, соответственно, неизвестно о существовании временного лага для переменной X , равного трем. Исходя из наблюдений, с помощью нейросети была проведена аппроксимация $Y(X)$ по значениям в точках наблюдений $t=1, 2, \dots, 40$. Так как и сама функция, и независимая переменная представлены временными рядами, необходимо использовать «историю» изменений, то есть формировать обучающий пример на основе наблюдений в нескольких временных точках. Исходя из рис. 2, ширина временного окна независимой переменной первоначально была принята равной двум, но после предварительных исследований это значение было увеличено до трех.

Нейросетевое моделирование проводилось с помощью пакета STATISTICA Neural Networks 4.0. Была построена нейронная сеть типа мно-

гослойный персептрон с одним скрытым слоем, логистической функцией активации структуры 4-4-1. Обучение проводилось методом обратного распространения ошибки.

В ходе анализа чувствительности нейросети была выдвинута гипотеза, что при наличии временного лага наибольшая ошибка предсказания будет наблюдаться при удалении того входа, который соответствует значению X в точке $t-\Delta$, которая в дальнейшем подтвердилась. Действительно, исходя из результатов анализа, проведенного средствами STATISTICA Neural Networks, представленных в табл. 1, наиболее значимым является вход, соответствующий $X(t-3)$. При сдвиге ряда X «назад» относительно Y на три временные точки, наиболее значимым становится $X(t)$, а точность прогнозирования увеличивается практически вдвое (структура сети и условия обучения оставались неизменными).

Таблица 1

Результаты анализа чувствительности нейросети для одной независимой переменной

	Переменная	Ранг	Чувствительность\абс.ошибка
До сдвига	$X(t)$	4	0.1591
	$X(t-1)$	3	0.1757
	$X(t-2)$	2	0.4127
	$X(t-3)$	1	0.8653
	Y		0.2373
После сдвига	$X(t)$	1	1.2168
	$X(t-1)$	2	0.1075
	$X(t-2)$	3	0.1067
	$X(t-3)$	4	0.0668
	Y		-0.1267

Для того, чтобы убедиться, что найденное значение временного лага является оптимальным, глубина погружения была увеличена на 1. Анализ чувствительности показал, что дополнительный вход $X(t-4)$ не является наиболее значимым.

Далее была рассмотрена функция двух переменных, имеющая вид $Y(t) = 10 \log(x_1^2(t-3) + 3x_2(t))$, график которой, как и зависимости $X_1(t)$ и $X_2(t)$ представлены на рис. 3. Иллюстрация анализа чувствительности нейросети структуры 8-17-1 средствами STATISTICA Neural Networks представлена на рис. 4. Сеть имеет 5 входов VAR1 – VAR5, на которые поступают значения X_1 (что соответствует ширине временного окна, равной четырем), и 3 входа VAR6 – VAR8 для отражения истории изменения переменной X_2 (ширина временного окна равна двум).

В ходе следующего вычислительного эксперимента были рассмотрены 8 независимых переменных ($X_1(t), \dots, X_8(t)$) и зависящая от них функция

$$Y(t) = 10 + \frac{(X_1(t) + \frac{X_2(t)}{3} - \ln(X_4)) \cdot X_3(t_3 - 1)}{X_6 + 2,5X_5(t_5 - 3)} - 0,44X_7(t_7 - 2) + X_8, \quad (3)$$

где $t_3 = t-1$, $t_5 = t-3$, $t_7 = t-2$.

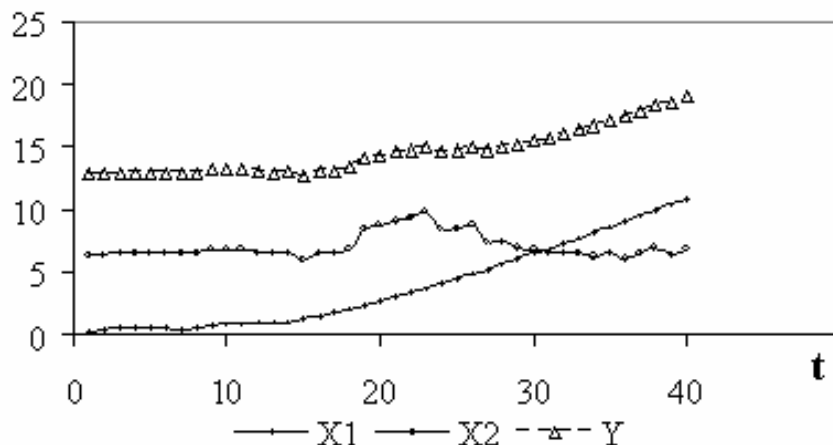


Рис. 3. Взаимное расположение временных рядов для двух независимых и одной зависимой переменных

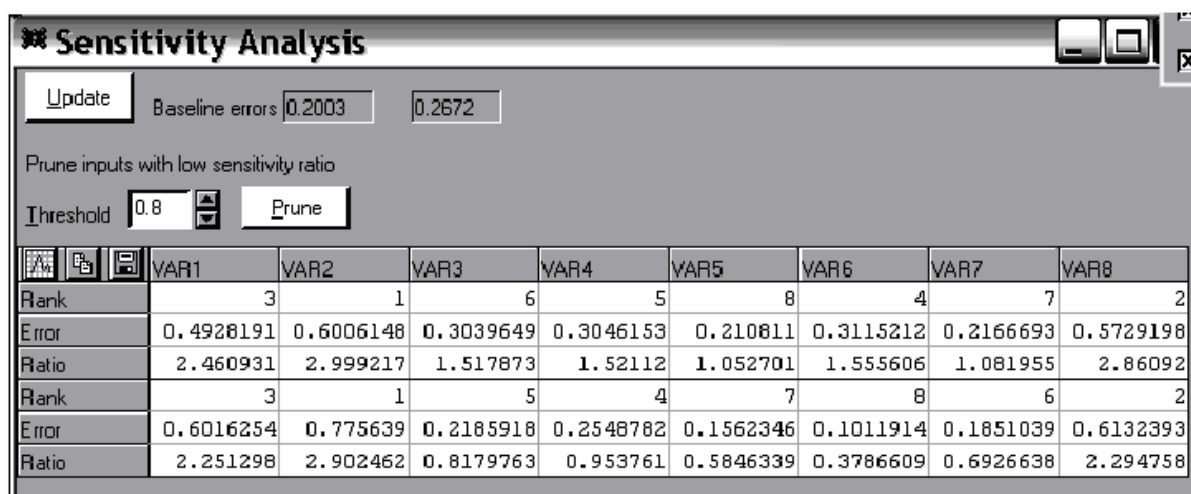


Рис. 4. Иллюстрация анализа чувствительности двух независимых переменных средствами STATISTICA Neural Networks

При увеличении количества учитываемых факторов и увеличении глубины погружения растет размерность задачи. Однако в случае коротких временных рядов имеющих наблюдения зачастую недостаточно для обучения сети с большим количеством входов и связей. Сеть «запоминает» обучающие примеры, о чем свидетельствует рост ошибки обобщения. Поэтому анализ чувствительности проводился по следующей схеме.

1 этап. Определение максимальной глубины погружения по графикам автокорреляционных функций независимых переменных и установление предельной ширины временного окна. Отмечено, что данная ширина не должна превышать половину длины временного ряда.

2 этап. Построение и обучение набора нейросетей с различным составом входов: глубина погружения исследуемого ряда принимается максимальной, остальные ряды представлены одним входом на ряд.

3 этап. Определение временного лага для каждого ряда с помощью анализа чувствительности.

Для исходных данных были построены графики автокорреляционных функций (рис. 5).

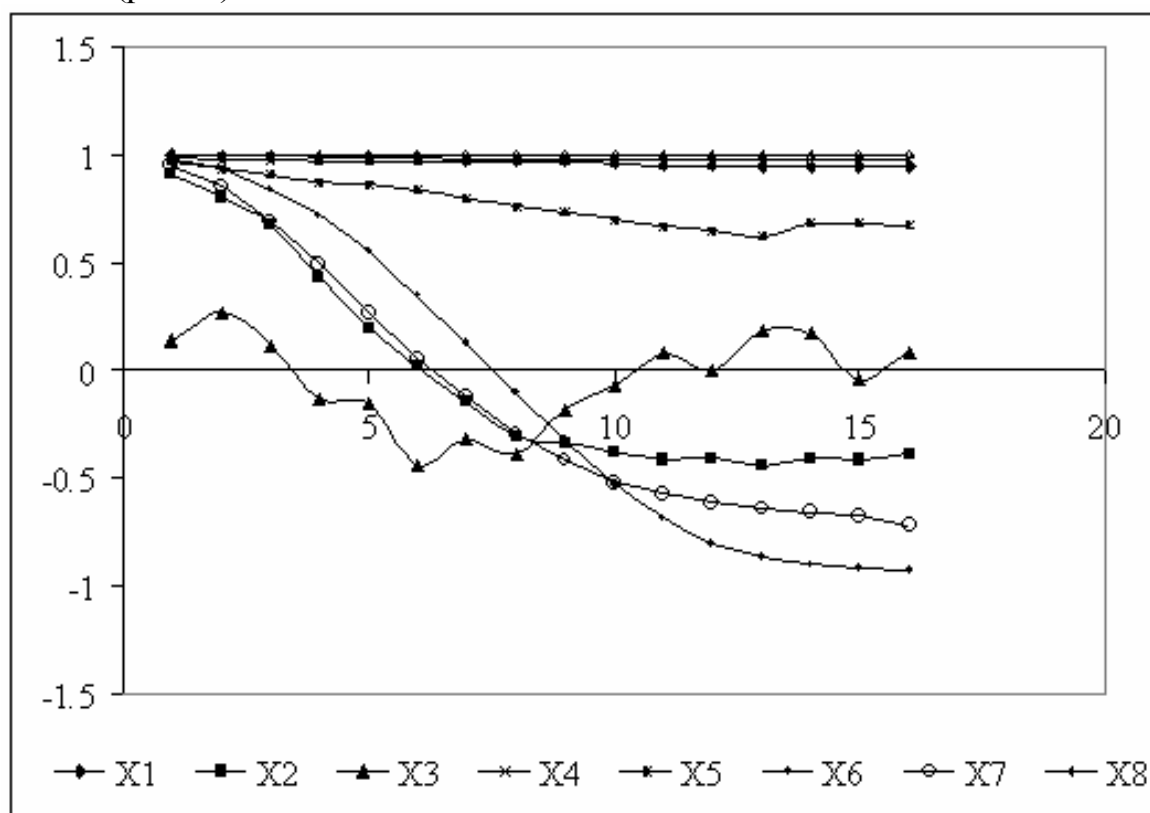


Рис. 5 - Графики автокорреляционных функций независимых переменных X_1, \dots, X_8

Исходя из полученных результатов анализа, максимальная ширина временного окна принята равной четырем.

Результаты анализа чувствительности набора нейросетей приведены на рис. 6. В рядах, изначально не обладавших временным лагом, наибольшая чувствительность отмечена для входов $X_i(t)$. Для рядов $X_3(t)$, $X_5(t)$, $X_7(t)$ наиболее чувствительными оказались входы $X_i(t-\Delta)$ ($\Delta_3=1$, $\Delta_5=3$, $\Delta_7=2$).

С учетом найденных временных лагов были сформирован смещенный массив исходных данных, использованный для обучения, тестирования и верификации нейросети. Нейросеть имела структуру 8-5-1, обучалась методом обратного распространения. Результаты обучения и прогнозирования на несмещенных и смещенных массивах представлены в таблице 2. Качество обучения оценивалось параметром S.D. Ratio – отношением стандартного отклонения ошибки к стандартному отклонению данных. Качество прогнозирования оценивалось показателем процентной погрешности MAPE [7], определяемым в виде

$$MAPE = \frac{|Y - Y^*|}{Y} \cdot 100\%, \quad (4)$$

где Y – фактическое значение выходной переменной, Y^* – спрогнозированное значение.

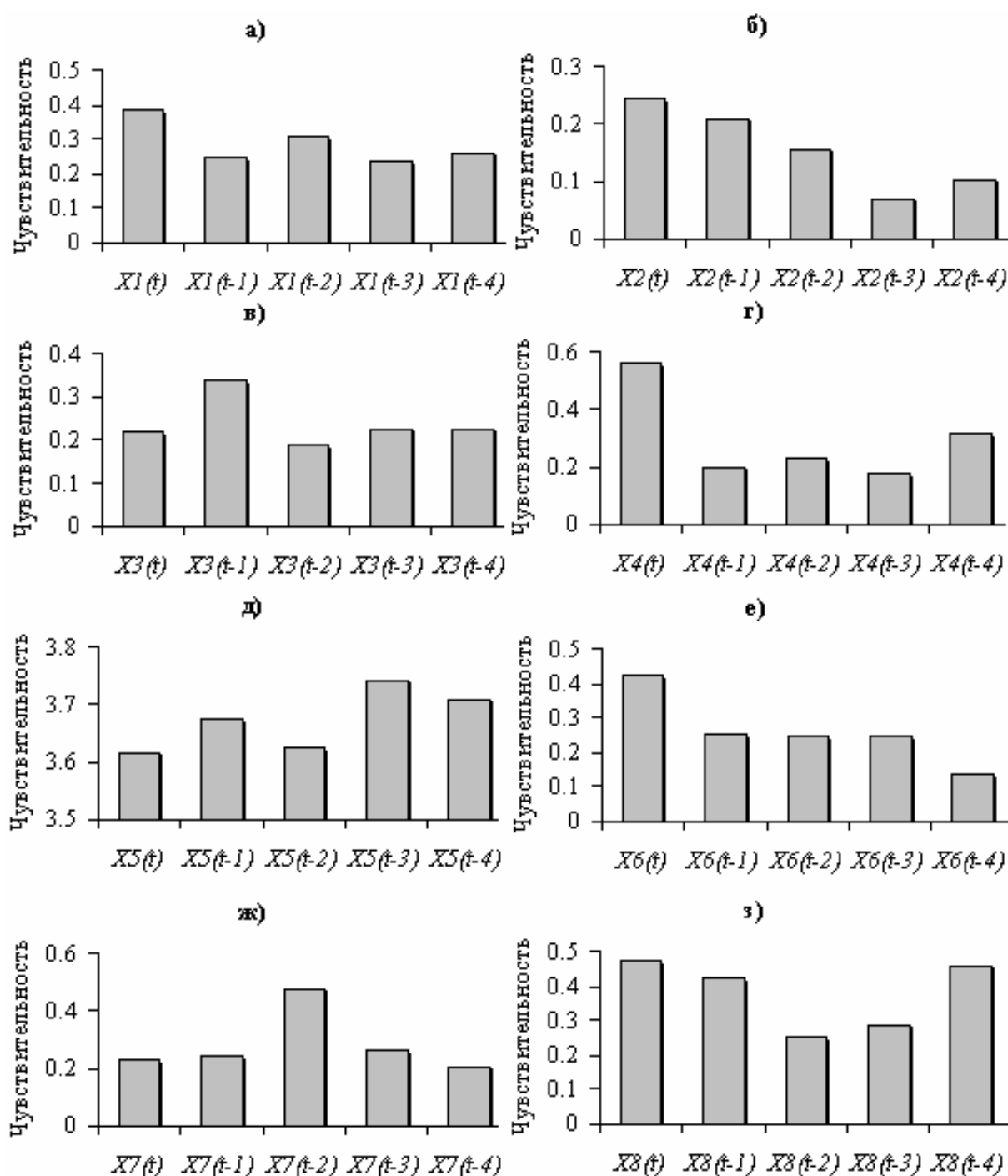


Рис. 6. Чувствительность входов, отражающих историю изменения независимой переменной

Как и ожидалось, качество прогноза на смещенных рядах значительно улучшилось.

Таким образом, нейросетевая модель чувствительна к наличию временного лага, характеризующего задержку влияния независимой переменной на прогнозируемую величину. Анализ чувствительности входов позволяет определить величину временного лага для каждой значимо влияющей переменной и построить уточненное множество исходных данных для обучения, тестирования и получения корректного прогноза.

Таблица 2

Качество обучения нейросети и прогноз Y (8 независимых переменных)

Набор входных переменных	S.D. Ratio (на обучающем множестве)	S.D. Ratio (на контрольном множестве)	MAPE, %
$X_1(t), \dots, X_8(t)$	0.240844	0.184005	2.78
$X_1(t), X_2(t), X_3(t-1), X_4(t), X_5(t-3), X_6(t), X_7(t-2), X_8(t)$	0.09965	0.199094	1.37

Список использованных источников

- 1.Кравец О.Я. Гибридные алгоритмы оптимизации моделей множественной регрессии на основе кросскорреляции. – Информационные технологии моделирования и управления, №4(22), 2005.
- 2.Крючкова И.Н., Кравец О.Я. Исследование особенностей нейросетевого моделирования социально-экономических процессов на смещенных временных рядах. – Информационные технологии моделирования и управления, №4(22), 2005.
- 3.Давыдов А. Социальное неблагополучие: зависимость объективных и субъективных оценок, НСН (<http://www.nns.ru/analit/pol/table.html>).
- 4.Родионов П.Е. Методика извлечения знаний в задачах анализа рядов динамики с использованием нейронных сетей. Автореферат диссертации на соискание ученой степени кандидата технических наук. – М., 2003.
- 5.Восьмиренко С.О. Разработка математического и программного обеспечения среды моделирования нейронных сетей для решения задач прогнозирования. Автореферат диссертации на соискание ученой степени кандидата технических наук. – М., 2005.
- 6.Уоссерман Ф. Нейрокомпьютерная техника: теория и практика. - М.: Мир, 1992.
- 7.Осовский С. Нейронные сети для обработки информации. – М.: Финансы и статистика, 2004. – 344 с.
- 8.Авдеева В.М., Кравец О.Я. Теоретические основы прогнозирования налоговых поступлений на основе кросскорреляционного анализа многомерных временных рядов// Системы управления и информационные технологии, 2006, №1.2(23). – С. 212-216.
- 9.Авдеева В.М., Кравец О.Я., Крючкова И.Н. Нейросетевое прогнозирование налоговых поступлений на смещенных временных рядах// Информационные технологии моделирования и управления. - 2006, №9(34), с. 1122-1130.

Журавлев С.В., Соломахин А.Н.

ОПТИМИЗАЦИЯ КАЛЕНДАРНЫХ СРОКОВ ПАРАЛЛЕЛЬНЫХ СТРОИТЕЛЬНЫХ РАБОТ В РАМКАХ СИСТЕМЫ МЕНЕДЖМЕНТА КРУПНОЙ ФИНАНСОВО-СТРОИТЕЛЬНОЙ КОРПОРАЦИИ СРЕДСТВАМИ ИНСТРУМЕНТАЛЬНЫХ КОМПОНЕНТ

*ЗАО Финансовая компания «Аксиома», г.Воронеж
Воронежский экономико-правовой институт*

1. Входные данные

При оптимизации календарных сроков параллельных строительных работ в рамках системы менеджмента крупной финансово-строительной корпорации используются следующие исходные данные для комплексного строительства (строительства множества объектов) в проекте организации строительства (ПОС) и проекте производства работ (ППР).