

**Министерство образования и науки Российской Федерации  
ФГАОУ ВО «Севастопольский государственный университет»**

**Институт информационных технологий  
и управления в технических системах**

**«Аналитическая система Deductor: Прогнозирование  
временных рядов»**

для студентов всех форм обучения направления подготовки

09.03.02 «Информационные системы и технологии»



Севастополь  
2017

**Цель:** научиться проводить поиск ассоциативных правил в Deductor Studio

**Время:** 2 часа

### Краткие теоретические сведения

**Временным рядом** (рядом динамики, динамическим рядом) называется последовательность значений показателя или признака, упорядоченная в хронологическом порядке, т.е. в порядке возрастания временного параметра. Отдельные наблюдения временного ряда называются **уровнями** этого ряда.

Каждый временной ряд содержит два элемента:

- 1) значения времени;
- 2) соответствующие им значения уровней ряда.

В качестве показателя времени в рядах динамики могут указываться либо определенные моменты времени (даты), либо отдельные периоды (сутки, месяцы, кварталы, полугодия, годы и т.д.). В зависимости от характера временного параметра ряды делятся на моментные и интервальные. В моментных рядах динамики уровни характеризуют значения показателя по состоянию на определенные моменты времени. Например, моментными являются временные ряды цен на определенные виды товаров, ряды курсов акций, уровни которых фиксируются для конкретных чисел. Примерами моментных рядов динамики могут служить также ряды численности населения или стоимости основных фондов, т.к. значения уровней этих рядов определяются ежегодно на одно и то же число. В интервальных рядах уровни характеризуют значение показателя за определенные интервалы (периоды) времени. Примерами могут служить ряды годовой (месячной, квартальной) динамики производства продукции в натуральном или стоимостном выражении.

Если уровни ряда представляют собой не непосредственно наблюдаемые значения, а производные величины (средние или относительные), то такие ряды называются производными. Уровни этих временных рядов получаются с помощью некоторых вычислений на основе абсолютных показателей.

Для успешного изучения динамики процесса важно, чтобы информация была полной, временной ряд имел достаточную длину (с учетом конкретных целей исследования). Например, при изучении периодических колебаний желательно иметь информацию не менее чем за три полных периода колебания. Поэтому при анализе сезонных колебаний на базе рядов месячной или квартальной динамики желательно иметь информацию, как правило, не менее чем за 3 года. Применение определенного математического аппарата также накладывает ограничение на допустимую длину временных рядов. Например, для использования регрессионного анализа требуется иметь временные ряды, длина которых в несколько раз превосходит количество независимых переменных. Уровни рядов динамики могут содержать аномальные значения или «выбросы». Часто появление таких значений может быть вызвано ошибками при сборе, записи и передаче информации. Выявление и исключение таких значений, замена их истинными или расчетными является необходимым этапом первичной обработки данных, т.к. применение математических методов к «засоренной» информации приводит к искажению результатов анализа. Однако аномальные значения могут отражать реальное развитие процесса, как, например, «скачок» курса доллара в «черный вторник».

В практике исследования динамики явлений и прогнозирования принято считать, что значения уровней временных рядов экономических показателей могут содержать следующие компоненты (составные части или структурно-образующие элементы):

- тренд;
- сезонную компоненту;
- циклическую компоненту;
- случайную составляющую.

Под **трендом** понимают изменение, определяющее общее направление развития, основную тенденцию временного ряда. Это систематическая составляющая долговременного действия. Наряду с долговременными тенденциями во временных рядах экономических процессов часто имеют место более или менее регулярные колебания – периодические составляющие рядов динамики. Если период колебаний не превышает одного года, то их называют **сезонными**. Чаще всего причиной их возникновения считаются природно-климатические условия. Примером могут служить колебания цен на сельскохозяйственную продукцию, в частности на картофель.

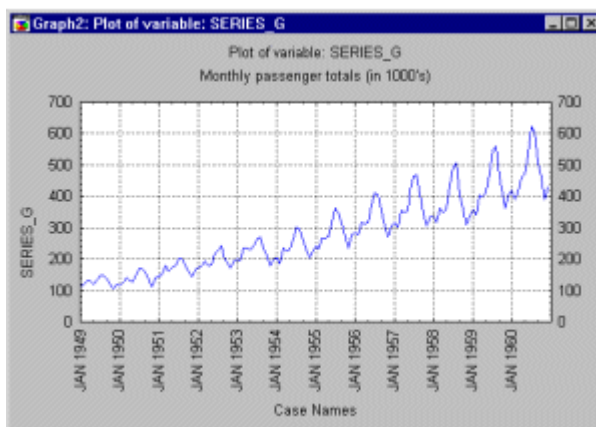
При большем периоде колебания считают, что во временных рядах имеет место **циклическая** составляющая. Примерами могут служить демографические, инвестиционные и другие циклы. Если из временного ряда удалить тренд и периодические составляющие, то останется **нерегулярная** компонента.

Разделяют факторы, под действием которых формируется нерегулярная компонента, на 2 вида:

- факторы резкого, внезапного действия;
- текущие факторы.

Факторы первого вида (например, стихийные бедствия, эпидемии и др.), как правило, вызывают более значительные отклонения. Иногда такие отклонения называют катастрофическими колебаниями. Факторы второго вида вызывают случайные колебания, являющиеся результатом действия большого числа побочных причин. Влияние каждого из текущих факторов незначительно, но ощущается их суммарное воздействие.

**Пример.**



Ряд G представляет месячные международные авиаперевозки (в тысячах) в течение 12 лет с 1949 по 1960. График месячных перевозок ясно показывает почти линейный тренд, т.е. имеется устойчивый рост перевозок из года в год (примерно в 4 раза больше пассажиров перевезено в 1960 году, чем в 1949). В то же время характер месячных перевозок повторяется, они имеют почти один и тот же характер в каждом годовом периоде (например, перевозок больше в отпускные периоды, чем в другие месяцы). Этот пример показывает довольно определенный тип модели временного ряда, в которой амплитуда сезонных изменений увеличивается вместе с трендом.

## Прогнозирование в Deductor Studio

Прогнозирование результата на определенное время вперед, основываясь на данных за прошедшее время, – задача, встречающаяся довольно часто. К примеру, перед большинством торговых фирм стоит задача оптимизации складских запасов, для решения которой требуется знать, что и сколько должно быть продано через неделю и т.п., задача предсказания стоимости

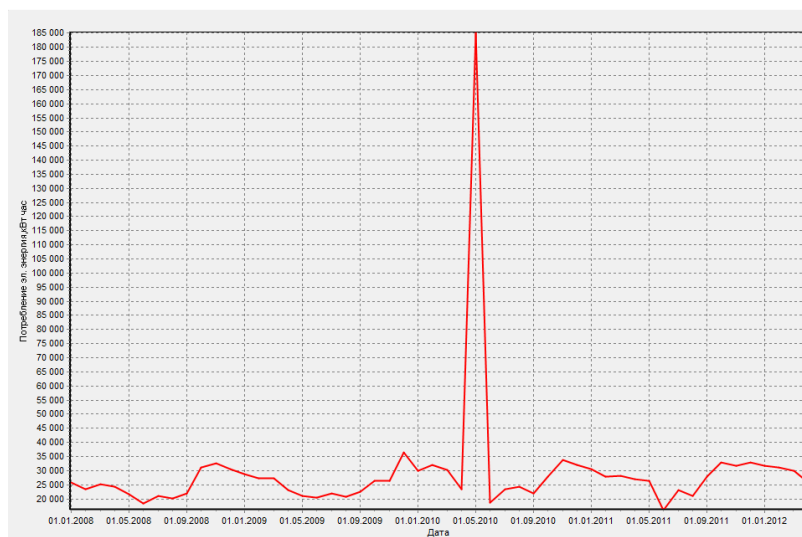
акций какого-нибудь предприятия через день и т.д. и другие подобные вопросы. Deductor Studio предлагает для этого инструмент "Прогнозирование".

Прогнозирование появляется в списке Мастера обработки только после построения какой-либо модели прогноза: нейросети, линейной регрессии и т.д. Прогнозировать на несколько шагов вперед имеет смысл только временной ряд (к примеру, если есть данные по недельным суммам продаж за определенный период, можно спрогнозировать сумму продаж на две недели вперед).

### Пример.

Исходные данные по потреблению электроэнергии находятся в файле "Потребление электрической энергии.txt".

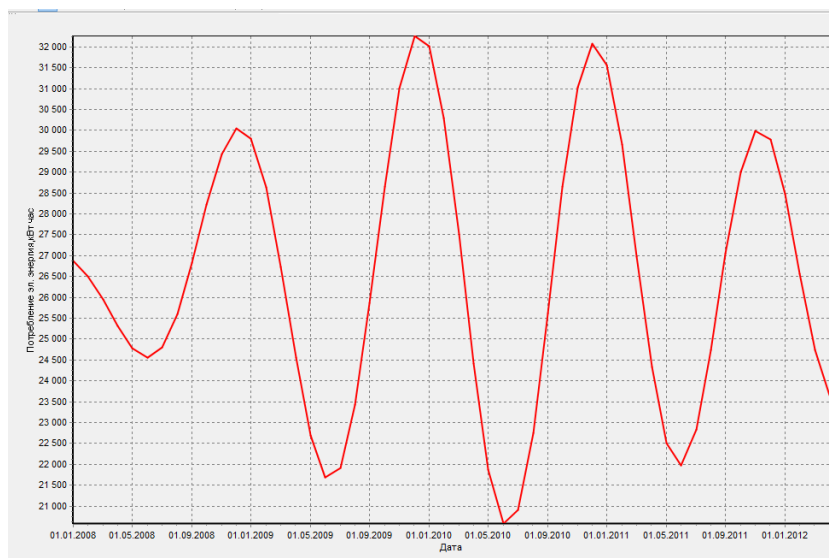
После импорта данных воспользуемся диаграммой для их просмотра.



На ней видно, что данные содержат аномалии (выбросы) и шумы, за которыми трудно разглядеть тенденцию. Поэтому перед прогнозированием необходимо удалить аномалии и сгладить данные.

Следующим шагом необходимо провести «Редактирование выбросов и экстремальных значений» и «Спектральную обработку»

Для просмотра данных можно воспользоваться диаграммой. Видно, что данные сгладились, аномалии и шумы исчезли.



Строить прогноз на будущее будем, основываясь на данных прошлых периодов, т. е. предполагая, что количество потребляемой энергии на следующий месяц зависит от количества потребляемой энергии за предыдущие месяцы. Это значит, что входными факторами для модели могут быть количество потребляемой энергии за текущий месяц, продажи за месяц ранее и т.д., а результатом должно быть количество потребляемой энергии за следующий месяц.

Для дальнейшей работы необходимо трансформировать данные к скользящему окну.

### Скользящее окно

При решении некоторых задач, например, при прогнозировании временных рядов при помощи нейросети, требуется подавать на вход модели значения нескольких смежных отсчетов из исходного набора данных. Такой метод отбора данных называется скользящим окном (окно – поскольку выделяется только некоторый непрерывный участок данных, скользящее – поскольку это окно «перемещается» по всему набору). При этом эффективность реализации заметно повышается, если не выбирать данные каждый раз из нескольких последовательных записей, а последовательно расположить данные, относящиеся к конкретной позиции окна, в одной записи.

Значения в одном из полей записи будут относиться к текущему отсчету, а в других – смещены от текущего отсчета «в будущее» или «в прошлое». Таким образом, преобразование скользящего окна имеет два параметра: «глубина погружения» - количество «прошлых» отсчетов, попадающих в окно, и «горизонт прогнозирования» – количество «будущих» отсчетов. Следует отметить, что для граничных (относительно начала и конца всей выборки) положений окна будут формироваться неполные записи, т.е. записи, содержащие пустые значения для отсутствующих прошлых или будущих отсчетов. Алгоритм преобразования позволяет исключить такие записи из выборки (тогда для нескольких граничных отсчетов записи формироваться не будут) либо включить их (тогда формируются записи для всех имеющихся отсчетов, но некоторые из них будут неполными). Для правильного формирования скользящего окна данные должны быть соответствующим образом упорядочены.

### Скользящее окно 12 месяцев назад

Запустить Мастер обработки, выбрать в качестве обработчика скользящее окно и перейти на следующий шаг.

Требуется выбрать глубину погружения 12, назначив поле "Количество" используемым. Тогда данные трансформируются к скользящему окну так, что аналитику будут доступны все нужные факторы для построения прогноза.

Дата	Потребление эл. энергия кВт час:12	Потребление эл. энергия кВт час:11	Потребление эл. энергия кВт час:10	Потребление эл. энергия кВт час:9
01.01.2009	26879,7916826241	26497,3238742571	25947,0310252638	25321,9720601086
01.02.2009	26497,3238742571	25947,0310252638	25321,9720601086	24792,945447299
01.03.2009	25947,0310252638	25321,9720601086	24792,945447299	24566,112737449
01.04.2009	25321,9720601086	24792,945447299	24566,112737449	24813,6544685508
01.05.2009	24792,945447299	24566,112737449	24813,6544685508	25599,24103298
01.06.2009	24566,112737449	24813,6544685508	25599,24103298	26826,0695167917
01.07.2009	24813,6544685508	25599,24103298	26826,0695167917	26231,4799139563
01.08.2009	25599,24103298	26826,0695167917	26231,4799139563	29438,9021035691
01.09.2009	26826,0695167917	26231,4799139563	29438,9021035691	30058,817094141
01.10.2009	26231,4799139563	29438,9021035691	30058,817094141	23811,7693287039
01.11.2009	29438,9021035691	30058,817094141	23811,7693287039	28634,9592624178
01.12.2009	30058,817094141	23811,7693287039	28634,9592624178	26734,4315637277
01.01.2010	23811,7693287039	28634,9592624178	26734,4315637277	24558,4655310325
01.02.2010	28634,9592624178	26734,4315637277	24558,4655310325	21691,1459590879
01.03.2010	26734,4315637277	24558,4655310325	21691,1459590879	21921,4007057107
01.04.2010	24558,4655310325	21691,1459590879	21921,4007057107	23422,1046728775
01.05.2010	21691,1459590879	21921,4007057107	23422,1046728775	25870,9257296058
01.06.2010	21921,4007057107	23422,1046728775	25870,9257296058	28649,7434237298
01.07.2010	23422,1046728775	25870,9257296058	28649,7434237298	31005,4236204159
01.08.2010	25870,9257296058	28649,7434237298	31005,4236204159	32261,7731788975
01.09.2010	28649,7434237298	31005,4236204159	32261,7731788975	32021,173884331
01.10.2010	31005,4236204159	32261,7731788975	32021,173884331	30294,0217076448
01.11.2010	32261,7731788975	32021,173884331	30294,0217076448	27513,1139417255
01.12.2010	32021,173884331	30294,0217076448	27513,1139417255	24423,4715263334
01.01.2011	30294,0217076448	27513,1139417255	24423,4715263334	21875,8798314448
01.02.2011	27513,1139417255	24423,4715263334	21875,8798314448	20582,9624529107
01.03.2011	24423,4715263334	21875,8798314448	20582,9624529107	20910,0045952944
01.04.2011	21875,8798314448	20582,9624529107	20910,0045952944	22764,2288484679
01.05.2011	20582,9624529107	20910,0045952944	22764,2288484679	25617,7875452123
01.06.2011	20910,0045952944	22764,2288484679	25617,7875452123	28659,7137325685
01.07.2011	22764,2288484679	25617,7875452123	28659,7137325685	31032,6809137336
01.08.2011	25617,7875452123	28659,7137325685	31032,6809137336	32083,9619308408
01.09.2011	28659,7137325685	31032,6809137336	32083,9619308408	31554,9353180312
01.10.2011	31032,6809137336	31554,9353180312	29651,7626259188	26976,0443268622
01.11.2011	31554,9353180312	29651,7626259188	26976,0443268622	24337,2368955682
01.12.2011	29651,7626259188	26976,0443268622	24337,2368955682	
01.01.2012	26976,0443268622			
01.02.2012	24337,2368955682			

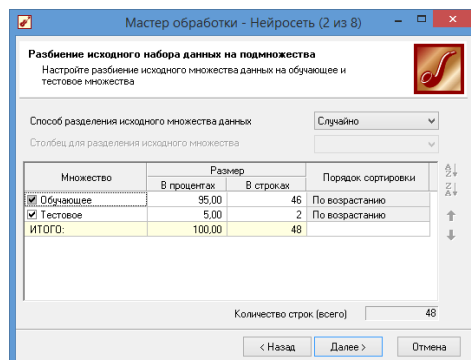
Теперь в качестве входных факторов можно использовать "Потребление эл. энергии - 12", "Потребление эл. энергии - 11" - данные по количеству потребляемой энергии 12 и 11 месяцев назад (относительно прогнозируемого месяца), и остальные. В качестве выходного поля можно использовать столбец "Потребление эл. энергии".

## Обучение нейросети

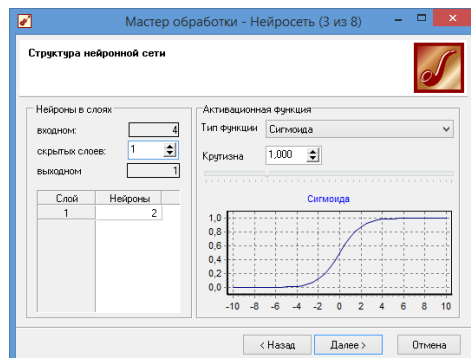
Построение модели прогноза.

Открыть Мастер обработки и выбрать в нем нейронную сеть. На втором шаге Мастера необходимо установить входные поля. Например можно взять "Потребление эл. энергии - 4", "Потребление эл. энергии - 3", "Потребление эл. энергии - 2" и "Потребление эл. энергии - 1" а в качестве выходного - "Потребление эл. энергии". Остальные поля сделать информационными.

На следующем шаге необходимо указать разбиение тестового и обучающего множеств.



На следующем этапе отмечаются необходимое количество слоев и нейронов в нейросети.

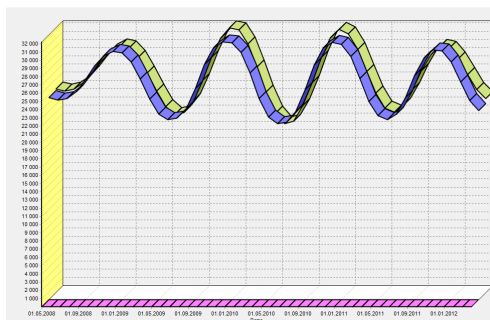


Перейдя далее, выбирается алгоритм обучения нейросети.

После построения модели для просмотра качества обучения полученные данные представить в виде диаграммы и диаграммы рассеяния.

В Мастере настройки диаграммы выбрать для отображения поля "Потребление эл. энергии" и "Потребление эл. энергии \_OUT" - реальное и спрогнозированное значение.

Результатом будет два графика.

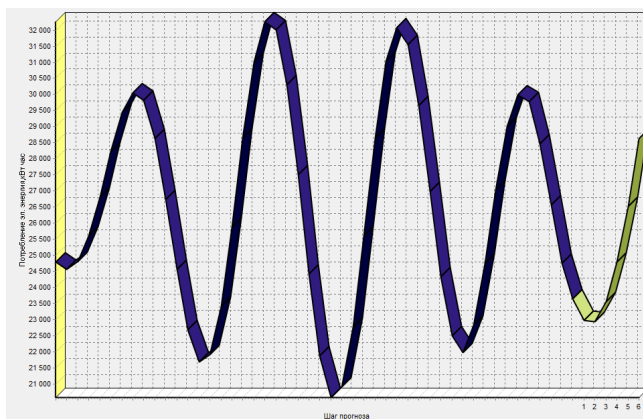


### Построение прогноза

Нейросеть обучена, осталось получить требуемый прогноз. Для этого необходимо открыть Мастер обработки и выбирать появившийся теперь обработчик "Прогнозирование".

На втором шаге Мастера предлагается настроить связи столбцов для прогнозирования временного ряда: откуда брать данные для столбца при очередном шаге прогноза.

Мастер сам верно настроил все переходы, поэтому остается только указать горизонт прогноза (на сколько вперед будем прогнозировать), а также для наглядности следует добавить к прогнозу исходные данные, установив в Мастере соответствующий флажок.



### Результат

После этого необходимо в качестве визуализатора выбрать "Диаграмму прогноза", которая появляется только после прогнозирования временного ряда.

В Мастере настройки столбцов диаграммы прогноза надо указать в качестве отображаемого столбец "Количество", а в качестве подписей по оси X указать столбец "Шаг прогноза".

Теперь аналитик может узнать количество потребленной электроэнергии в следующем месяце и шесть месяцев спустя.

### Задание и порядок выполнения практической работы №6

#### Задание 1:

Подготовить данные для прогнозирования временного ряда.

#### Задание 2:

Провести анализ временного ряда.

#### Задание 3:

Провести прогнозирование временного ряда на 6 месяцев.