

ЭКСПЕРТНО-СТАТИСТИЧЕСКИЕ СИСТЕМЫ ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ И ИМИТАЦИОННО-ОЦЕНОЧНОЕ МОДЕЛИРОВАНИЕ¹

А.Г. Беляков, А.С. Мандель, Н.И. Борзенко, А.В. Лапин, А.Б. Токмакова, В.В. Шушков

Институт проблем управления им. В.А. Трапезникова, г. Москва

Описаны процедуры прогнозирования на основе метода аналогов. Рассмотрены две программные системы, реализующие совокупность экспертно-статистических процедур прогнозирования временных рядов по коротким выборкам (система ЭКСПАМ) и схему имитационного моделирования процесса экспертно-статистических систем прогнозирования на стадии предварительного обследования объекта прогнозирования (система ЭКСПРИМ).

ВВЕДЕНИЕ

При прогнозировании различного рода процессов в экономике и технике нередки ситуации, когда сведения о поведении объекта в его предыстории либо весьма ограничены, либо — даже если они достаточно полны — имеют одну особенность, которая заключается в том, что к моменту выполнения прогнозных расчетов тенденции развития процесса успевают претерпеть серьёзные изменения. Не исключён также вариант острой необходимости прогнозирования будущих значений временного ряда в те моменты времени, когда известны только самые первые значения ряда либо выборка вообще пуста. Например, в случае выхода нового товара на рынок, когда успех или провал обусловлен первыми неделями продаж. В силу специфики прикладной проблемы промедление в формировании прогноза может впоследствии сделать саму процедуру прогнозирования неактуальной. В подобных условиях традиционные статистические методы прогнозирования зачастую не работают вообще либо приводят к недостоверным результатам. В результате приходится прибегать к использованию дополнительной, чаще всего, экспертной информации.

Один из способов прогнозирования с использованием экспертной информации состоит в прогнозировании на основе подбора аналогов. Суть метода аналогов связана с гипотезой о том, что в подобных ситуациях опытные эксперты-предметники обычно обращаются к своему профессиональному багажу и стараются сформировать прогноз, опираясь на “прецеденты”, т. е. на наблюдавшиеся ранее похожие процессы или объекты — аналоги. Для многих предметных областей выполнение этой гипотезы отлично подтверждается практикой.

Метод аналогов можно применять в ситуациях, когда:

- отсутствует или крайне ограничена статистическая информация об объекте прогнозирования, и прогнозирование можно осуществить, руководствуясь, прежде всего, информацией субъективного характера;
- по каким-либо причинам эксперты не желают или затрудняются сформировать интервальные или точечные оценки будущих значений временного ряда;
- наличествует экспертная информация об объекте прогнозирования, позволяющая отнести его к тому или иному классу объектов;
- в распоряжении экспертов имеется представительный набор статистических данных о длинных отрезках предыстории для значительного числа ранее наблюдавшихся объектов, которые могут служить потенциальными аналогами нового объекта.

¹ Статья рекомендована к печати Программным комитетом Второй международной конференции по проблемам управления (Москва, 2003 г.).



Процедура прогнозирования на основе метода аналогов обладает двумя серьёзными преимуществами:

- диалог с экспертом ведётся на хорошо понятном ему языке (из его “родной” предметной области), без употребления специальных статистических терминов типа “тренд”, “значения временных рядов” и т. п;
- процесс подбора аналогов позволяет собирать большие объёмы экспертной информации об объекте прогнозирования; при этом вполне возможно не только построение тренда на основе временных рядов объектов-аналогов или вычисление интервальных оценок значений конкретных членов временных рядов, но и применение других методов статистической обработки информации об объектах-аналогах.

КРАТКИЙ ОЧЕРК МЕТОДА АНАЛОГОВ [1]

Предъявляемый эксперту объект прогнозирования (ОП) представляет собой отрезок временного ряда длины N : $y(n)$, $n = 0, 1, 2, \dots, N$ (в частном случае N может быть равно нулю: полное отсутствие выборки данных). В ответ на предъявленную выборку данных (и/или сообщаемое ему наименование ОП) эксперт формирует перечень объектов-аналогов, которые представлены в базе данных системы прогнозирования “полными” временными рядами, т. е. рядами, длина которых существенно превосходит число N .

Пусть Z – множество номеров отобранных экспертом объектов-аналогов, для каждого из которых эксперт имеет право (но не обязан) задать две числовые характеристики: *коэффициент похожести* l_k , $k \in Z$, и *коэффициент масштаба* s_k , $k \in Z$. В данном контексте эксперт самостоятельно (без каких-либо комментариев со стороны разработчика) определяет смысл понятия “похожесть”, характеризуя тем самым свои неформализованные, чисто субъективные представления о качестве подобранного им аналога в смысле его сходства с ОП. Что касается коэффициента масштаба, то его отличное от единицы значение означает, что временной ряд объекта-аналога по форме близок к временному ряду ОП, но отличается от него размахом. Если эксперт затрудняется или “ленится” дать оценки соответствующих коэффициентов, то по умолчанию они полагаются равными единице.

По всем потенциальным аналогам база данных экспертно-статистической системы (ЭСС) прогнозирования содержит информацию о временных рядах процессов функционирования объектов-аналогов в форме зафиксированных в базе данных реализаций $\{x_k(n), n = 0, 1, 2, \dots, N_1\}$, где $N_1 \gg N$. После того, как множество аналогов Z выбрано,

предметом изучения становится набор реализаций временных рядов аналогов $\{x_k(n), k \in Z, n = 0, 1, 2, \dots, N_1\}$, а также множества значений коэффициентов похожести $\{l_k, k \in Z\}$ и множества коэффициентов масштаба $\{s_k, k \in Z\}$.

Для вычисления прогноза значений временного ряда ОП в моменты времени n , $n > N$, используется следующая формула:

$$y(n) = \frac{1}{L} \sum_{k \in Z} \alpha_k l_k s_k x(n), \quad (1)$$

где $L = \sum_{k \in Z} l_k$. При $N > 0$ значения коэффициентов α_k , $k \in Z$, определяются в результате решения следующей задачи минимизации:

$$\min_{\{\alpha_k, k \in Z\}} \sum_{n=1}^N \left(\frac{1}{L} \sum_{k \in Z} \alpha_k l_k s_k x(n) - y(n) \right)^2. \quad (2)$$

Смысл этой задачи заключается в том, что в результате её решения достигается оптимальная интерполяция уже наблюдавшаяся в интервале длины N части реализации временного ряда ОП $y(n)$, $n = 0, 1, 2, \dots, N$ набором дискретных по времени кривых $\{x_k(n), k \in Z, n = 0, 1, 2, \dots, N\}$.

Если $N = 0$ (т. е. выборка данных по ОП вообще отсутствует), то все α_k , $k \in Z$, полагаются равными единице.

ВНЕСЕНИЕ КОРРЕКТИРОВОК В ПРОЦЕСС ПРОГНОЗИРОВАНИЯ

Если в результате применения методов автоматической классификации [2] вся информация по объектам, хранящимся в базе данных ЭСС прогнозирования, структурирована, то для уточнения формируемых системой и экспертом прогнозов целесообразна организация диалога между экспертом и ЭСС с целью корректировки множества объектов-аналогов.

Действительно, в рамках метода аналогов точность прогноза зависит только от правильности подбора множества аналогов (и, разумеется, от квалификации эксперта-прогнозиста). Следует помнить, что результаты автоматической классификации и той неявной классификации, которую мы несколько вольно назовём “практическим опытом эксперта”, могут не совпадать, а эксперт обладает правом “дезаурировать” любые рекомендации ЭСС. При формировании классов похожести эксперт и ЭСС могут пользоваться разными “метриками”. Таким образом, в процессе диалога с целью корректировки множества аналогов ЭСС работает только в режиме советчика.

Интерактивная процедура взаимодействия ЭСС и эксперта базируется на задании упорядоченного набора метрических параметров $\{d_1, d_2, \dots, d_R\}$, $d_1 < d_2 < \dots < d_R$, представляющих собой пороговые значения расстояний между построенными в результате применения процедур автоматической классификации классами. По этим порогам осуществляется экспертно-статистическая оценка “качества” выбранного экспертом множества аналогов.

Пусть экспертом отобрано множество аналогов Z и пусть в результате решения задачи автоматической классификации все объекты базы данных отнесены к некоторым классам похожести A_i , $i = 1, 2, \dots, M$. Возможны следующие основные варианты.

- Все объекты множества Z принадлежат одному из классов или нескольким классам похожести, “центры” которых находятся в пределах сферы диаметром d_1 . В этом случае ЭСС “молчит”, признавая выбор эксперта непротиворечивым, и никаких “советов” по корректировке множества Z эксперту не предлагает.
- Большая часть объектов множества Z принадлежит одному из классов или нескольким классам похожести, “центры” которых находятся в пределах сферы диаметром d_1 . Однако имеется и некоторое количество объектов, которые выходят за пределы шара d_1 и попадают в шары диаметрами d_2, \dots, d_R . В этом случае ЭСС информирует эксперта о некотором расхождении между объектами, используя формулировки, “сдержанность” которых зависит от степени удаления непохожих объектов (от “мне кажется” до “настоятельно рекомендую пересмотреть...”).
- Нет ни единой массовой группы объектов из множества Z , которые принадлежали бы одному-единственному шару диаметром d_1 . В этом случае определяются максимально удаленные между собой объекты, которые и выводятся на показ эксперту (например, отрезками соответствующих временных рядов) с информированием эксперта о том, что в осуществленной им выборке объектов-аналогов имеются заметные расхождения. Уровень существенности этих расхождений, как и выше, характеризуется с различной степенью “жесткости” (см. предыдущий пункт) в зависимости от того, какой из порогов d_2, \dots, d_R оказался превышенным.

При выборе аналогов помимо корректировки самого множества аналогов эксперту необходимо предоставить возможность внесения различных количественных поправок. Укажем некоторые варианты подобных корректировок: внесение изменений в значения весовых коэффициентов, которые характеризуют степень близости прогнози-

руемого ряда (ОП) к тем или иным аналогам из выбранного экспертом списка; непосредственные корректировки рекомендуемых системой поддержки принятия решений будущих значений временного ряда для ОП (прогнозов).

Таким образом, первоначальный прогноз формируется на основе выбранных экспертом аналогов с учётом заданных заранее весовых коэффициентов. Если в прогнозировании принимают участие несколько экспертов, то в качестве весовых коэффициентов может выступать частота упоминания экспертами каждого из аналогов.

ЭКСПАМ – ЭКСПЕРТНО-СТАТИСТИЧЕСКАЯ СИСТЕМА ПРОГНОЗИРОВАНИЯ ПО МЕТОДУ АНАЛОГОВ

Метод аналогов реализован в программной системе ЭКСПАМ, предназначенной для выполнения экспертно-статистической обработки при прогнозировании временных рядов.

Система ЭКСПАМ ориентирована на тех экспертов-предметников, для которых одна из главных задач состоит в прогнозировании развития процессов и принятии решений на основе разнородной информации. Поскольку потенциальными пользователями системы оказываются эксперты по различным предметным областям, то задача разработчиков заключалась в создании простого “дружественного” интерфейса, который, с одной стороны, был бы достаточно универсальным, т. е. мог быть применен в максимально возможном числе предметных областей, с другой – легко настраиваемым (под терминологический базис конкретной предметной области).

Как сказано выше, пользователями системы ЭКСПАМ являются эксперты, занимающиеся прогнозированием развития различных явлений, описывающихся временными рядами. Учитывая распространенную на практике и вполне естественную специфику процедур принятия решений, эксперты-пользователи системы делятся на две группы: “рядовые” эксперты, формирующие прогнозы на основе своего профессионального опыта и подготовленной для них статистической информации, и “главный” эксперт (эксперты), имеющий возможность просматривать и анализировать прогнозы “рядовых” экспертов, и на их основе принимать окончательные решения.

Структура системы ЭКСПАМ

Система состоит из следующих основных разделов:

- раздел предварительного прогнозирования;
- раздел прогнозирования на основе имеющихся статистических данных;
- раздел главного эксперта;
- раздел аналитики по экспертам.

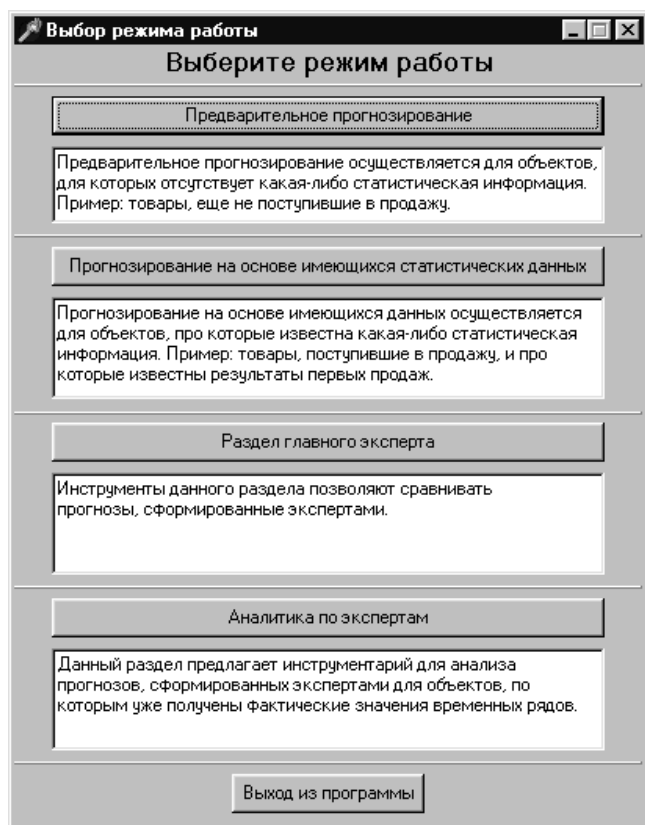


Рис. 1. Окно выбора разделов системы ЭКСПАМ

Раздел предварительного прогнозирования предназначен для формирования прогнозов временных рядов в те моменты времени, когда для объектов прогнозирования ещё отсутствуют результаты наблюдения описывающих их временных рядов. Основным инструментом данного раздела — прогнозирование на основе метода аналогов в чистом виде. С помощью интерфейса раздела эксперты выбирают объекты, которые, по их мнению, похожи на ОП. После ввода ряда коэффициентов и просмотра предварительных прогнозов работающий с системой эксперт утверждает список аналогов и сформированный на их основе прогноз. Результаты предварительного прогнозирования, списки аналогов, а также другие виды информации о прогнозе сохраняются в базе данных системы.

Раздел прогнозирования на основе имеющихся статистических данных включает в свой состав инструменты для формирования прогнозов в те моменты времени, когда для ОП уже получены первые наблюдавшиеся значения временного ряда. В данном разделе эксперты имеют возможность просматривать ранее сформированные прогнозы, изучать отклонения фактически полученных значений ряда от ранее спрогнозированных, знакомится с новыми прогнозами, сформированными системой на

основе модели авторегрессии, корректировать списки аналогов и новые прогнозы.

В *разделе главного эксперта* лицо, принимающее решения, — в дальнейшем мы будем называть его *главным экспертом* — может изучать и анализировать прогнозы и списки аналогов, сформированные “рядовыми” экспертами, и на их основе строить собственный, окончательный прогноз временного ряда. В данном разделе формируются “коэффициенты доверия к рядовым экспертам”, которыми описывается степень доверия главного эксперта к прогнозам рядовых экспертов. В дальнейшем коэффициенты доверия выступают в роли весовых коэффициентов при комбинировании прогнозов различных экспертов.

В *разделе аналитики по экспертам* содержится набор средств, позволяющих просматривать и анализировать историю формирования прогнозов и статистические данные по прогнозам, построенных рядовыми экспертами-пользователями системы. Раздел предназначен для главного эксперта и позволяет сравнивать прогнозы различных рядовых экспертов, сформированные для одних и тех же ОП, просматривать статистическую информацию об ошибках прогнозов каждого эксперта. Используется набор специально разработанных алгоритмов, которые позволяют так обработать предысторию формирования прогнозов рядовыми экспертами, чтобы оценить достоверность их действий и рекомендовать главному эксперту установить определённые значения коэффициентов доверия.

Окно выбора разделов представлено на рис. 1.

Рассмотрим некоторые детали процедур, реализованных в каждом из упомянутых разделов.

В разделе предварительного прогнозирования при отборе аналогов пользователь системы может выполнять следующие действия:

- просматривать статистическую информацию об объектах из базы данных, а также временные ряды в графическом и табличном представлениях;
- для каждого выбранного объекта вводить ряд коэффициентов, которые используются при формировании прогнозов, а именно:
 - коэффициент масштаба s_k , где k — номер аналога, описывающий соотношение между значениями элементов временного ряда объекта прогнозирования и значениями временного ряда объекта-аналога;
 - коэффициент похожести l_k , где k — номер аналога, описывающий степень уверенности эксперта в том, что временной ряд объекта прогнозирования будет близок к временному ряду объекта-аналога (с учетом коэффициента масштаба).

По умолчанию значение коэффициента масштаба устанавливается равным единице, а значе-

ние коэффициента похожести – 100% (в системе ЭКСПАМ оказалось удобнее использовать процентные, а не абсолютные, как в формуле (1), значения коэффициентов l_k).

После того, как экспертом выбраны все аналоги и заданы все коэффициенты, система строит прогноз. Значения прогнозируемого временного ряда вычисляются по упрощенному варианту формулы (1):

$$f_j = \frac{1}{L} \cdot \sum_{k=1}^K l_k s_k A_{kj}, \quad (3)$$

где K – количество аналогов, отобранных экспертом для объекта прогнозирования; f_j – значение

прогноза в точке j ; $L = \sum_{k=1}^K l_k$ – сумма значений ко-

эффициентов похожести для всех аналогов, отобранных для данного объекта; A_{kj} – значение временного ряда k -го аналога в j -й точке. В системе ЭКСПАМ процедура минимизации по коэффициентам α_k (см. формулу (2)) отсутствует, и все значения α_k полагаются равными единице.

Длина прогнозного периода (или, иными словами, длина периода упреждения прогноза), т. е. число точек, для которых осуществляется прогнозирование, задаётся пользователем.

Сформированный прогноз предлагается пользователю в виде графика и таблицы значений.

Пользователь системы имеет возможность скорректировать предложенный системой прогноз одним из следующих способов:

- скорректировать список аналогов;
- изменить значения введённых коэффициентов масштаба и похожести;
- вручную изменить значения прогнозного ряда.

В результате выполнения описанных шагов в базу данных системы заносятся:

- предварительный прогноз эксперта с необходимыми комментариями (имя эксперта, дата формирования прогноза и др.);
- список выбранных аналогов и введённых коэффициентов похожести и масштаба;
- значения ручных корректировок, осуществлённых экспертом применительно к прогнозу, который сформирован системой на основе списка аналогов, и соответствующих коэффициентов.

Раздел прогнозирования на основе имеющихся статистических данных, как следует из его названия, предназначен для формирования новых или корректировки ранее построенных прогнозов в те моменты времени, когда для объекта прогнозирования уже известны первые члены описывающего его временного ряда.

Главные задачи раздела:

- предоставление экспертам подробной информации о расхождениях между значениями сформированного ранее предварительного прогноза и фактически полученными значениями временного ряда;
- корректировка прогноза с учётом полученных статистических данных;
- построение прогнозов с помощью статистических методов;
- построение прогнозов на основе методов, объединяющих данные объективного и субъективного характеров;
- предоставление средств, позволяющих эксперту сравнивать, комбинировать и корректировать имеющиеся прогнозы с целью формирования нового, более точного прогноза.

После выбора пользователем объекта, для которого желательно уточнить прогноз, система предъявляет текущую информацию об уже известных членах ряда и об их расхождении с построенным прогнозом. Пользователю предоставляется следующая информация:

- отклонения прогнозного ряда от полученных фактических значений ряда для каждого уже известного члена ряда; эти отклонения демонстрируются в форме абсолютных значений и в виде процентов от фактических значений ряда;
- графики прогноза и статистики и график ряда ошибок прогноза;
- среднее значение, выборочная дисперсия и выборочное среднеквадратическое отклонение ряда ошибок прогноза;
- суммарные значения ошибки прогноза по всем известным точкам с учётом и без учёта знаков ошибок по каждой точке.

После просмотра информации пользователь может обратиться к встроенному в данный раздел системы пакету программ, который формирует прогнозы на базе модели авторегрессии.

Выбор оптимальной модели включает в себя следующие шаги:

- задание максимального порядка модели (в качестве потенциальных моделей рассматриваются только те модели, порядок которых не превышает заданную величину; по умолчанию максимальный порядок равен пяти, минимальный – единице);
- построение моделей авторегрессии всех порядков от минимального до максимального; значения параметров каждой модели вычисляются с помощью решения матричного уравнения Юла–Уокера:

$$AB = F,$$



где

$$A = \begin{pmatrix} R(0) & R(1) & \dots & R(p-1) \\ P(1) & 1 & & R(p-2) \\ \dots & & & \\ R(p-1) & R(p-2) & \dots & R(0) \end{pmatrix};$$

$$F = \begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_p \end{pmatrix}; \quad B = \begin{pmatrix} R(1) \\ R(2) \\ \dots \\ R(p) \end{pmatrix};$$

$$R(k) = \frac{1}{N} \sum_{i=1}^{N-k} (x_i - \bar{x})(x_{i+k} - \bar{x}) - \text{выборочная ав-}$$

токорреляционная функция; \bar{x} – выборочное среднее;

– выбор оптимального по критерию Акаике порядка модели авторегрессии:

$$\bar{P} = \arg \min_{p_{\min} \leq p \leq p_{\max}} (N \ln(\text{var}(u) + 2p),$$

где \bar{P} , p_{\min} и p_{\max} – соответственно оптимальный, минимальный и максимальный порядки модели; $\text{var}(u)$ – дисперсия остатков модели, вычисляемая по формуле $\text{var}(u) = \text{var}(x)(1 - a_1 r(1) - \dots - a_p r(p)$,

где $r(k) = \frac{R(k)}{R(0)}$ – нормированная выборочная ав-токорреляционная функция; N – объем выборки.

В разделе прогнозирования на основе имеющихся статистических данных также реализован автоматический поиск объектов-аналогов, близкий по форме к процедурам поиска из раздела предварительного прогнозирования. Однако, в отличие от раздела предварительного прогнозирования, в данном случае ведётся приоритетный поиск тех объектов, значения временных рядов которых близки к фактически полученным значениям временного ряда, описывающего объект прогнозирования. Только после этого учитывается степень близости этих точек к членам прогнозного ряда, сформированного самим экспертом.

Для решения данной задачи вводится новая метрика r , а именно:

- задаётся число W , обозначающее количество точек, которые принимаются во внимание при поиске объектов-аналогов (по умолчанию $W = 10$);
- расстояние между объектами вычисляется по формуле:

$$r_{lm}^s = \gamma \frac{\sum_{i=1}^{W_1} (x_{li} - x_{mi})^2}{\sum_{i=1}^{W_1} x_{li}^2} + \delta \frac{\sum_{i=W_1+1}^W (x_{li} - x_{mi})^2}{\sum_{i=W_1+1}^W x_{li}^2},$$

где x_{li} и x_{mi} – i -е значение l -го и m -го временного ряда соответственно; γ и δ – коэффициенты, определяющие соответственно вклад известных статистических данных и прогнозируемых данных в расстояние между объектами, W_1 – число уже известных членов временного ряда.

Раздел главного эксперта предназначен для построения окончательного прогноза на основе прогнозов, сформированных рядовыми экспертами. С помощью средств данного раздела главный эксперт имеет возможность сравнивать прогнозы, сформированные для данного объекта другими экспертами, и с помощью ввода и корректировки коэффициентов доверия к экспертам сближать окончательный прогноз с прогнозами наиболее грамотных, с его точки зрения, рядовых экспертов. Коэффициент доверия к эксперту принимает значение от 0 до 100%. Окончательный прогноз вычисляется по следующей формуле:

$$x_i = \frac{1}{S_{\text{КД}}} \sum_{j=1}^M K_{\text{Д}j} x_{ij},$$

где x_i – значение i -й точки прогнозного ряда; $S_{\text{КД}}$ – сумма коэффициентов доверия по всем экспертам, сформировавшим прогнозы для данного объекта; M – число экспертов, сформировавших прогнозы для данного объекта; $K_{\text{Д}j}$ – коэффициент доверия j -му эксперту; x_{ij} – значение i -й точки в прогнозе, сформированном j -м экспертом.

В разделе аналитики по экспертам содержится разнообразная информация о прогнозах, сформированных рядовыми экспертами. Основной задачей данного раздела является предоставление главному эксперту информации и рекомендаций, на основе которых он мог бы делать выводы о точности прогнозов, формируемых рядовыми экспертами-пользователями системы.

Раздел состоит из трёх подразделов:

- аналитика по прогнозам, сформированным каждым рядовым экспертом;
- сравнительный анализ прогнозов, сформированных экспертами;
- общая аналитика по всем экспертам.

Описание экспериментов с системой ЭКСПАМ

Ниже приводится описание одного из натуральных экспериментов с системой ЭКСПАМ, когда перед экспертами компании-производителя, выпускающей широкий ассортимент продукции, была поставлена задача прогнозирования спроса на новый вид продукции. На базе прогноза, утвержденного главным экспертом компании, принималось решение об объёме производства этого нового вида. Главный эксперт принимал решение, руководствуясь своим профессиональным опытом и интуици-

ей, а также на основе прогнозов, подготовленных несколькими рядовыми экспертами компании независимо друг от друга.

Новый вид продукции будем называть “новым объектом”. Спрос на него представляется в виде временного ряда, каждый член которого описывает число проданных за одну неделю экземпляров нового объекта. Специфика рассматриваемой предметной области заключается в том, что спрос на новую продукцию, как правило, резко падает после первых недель присутствия товара на рынке (товары-“хиты”), и ошибки в предварительных прогнозах могут приводить к крайне неприятным экономическим последствиям для компании-производителя.

По каждому наименованию видов продукции, поступающих (или поступавших ранее) на рынок, в базе данных системы содержится информация следующих двух типов.

- Статистическая информация, т. е. временные ряды, описывающие зафиксированный ранее спрос на данный вид продукции, и статистические характеристики соответствующих рядов.
- Экспертная информация, т. е. различные характеристики и сведения об объектах, принадлежащих данной предметной области. Объекты из каждой предметной области обладают более или менее фиксированным и чётко определенным набором характеристик, поэтому после нескольких сеансов общения с экспертами, как правило, не составляет большого труда описать этот набор на языке разработчиков. В результате вся информация в базе данных системы оказывается структурированной, а это позволяет достаточно просто разбивать объекты на классы похожести и осуществлять другие необходимые в процессе прогнозирования операции.

Таким образом, перед экспертами была поставлена задача прогнозирования спроса на новый объект. Зафиксированный в результате спрос на новый объект показан на рис. 2. По оси абсцисс отложены номера недель, в которые фиксировался спрос на новый объект, а по оси ординат – уровень спроса в некоторых условных единицах.

Первый неформальный шаг процесса прогнозирования заключался в том, что эксперты компании на основе известной информации и характеристиках нового вида продукции причислили его к тому или иному классу. Важно отметить, что, руководствуясь известными характеристиками товара и (или) пользуясь собственными (грубыми) оценками спроса на новый вид продукции в первые две недели, эксперт может резко сузить список потенциальных объектов-аналогов.

Следующий шаг состоял в выборе аналогов в базе данных системы и вводе коэффициентов масштаба и похожести для отобранных аналогов. На основе известных характеристик нового объекта

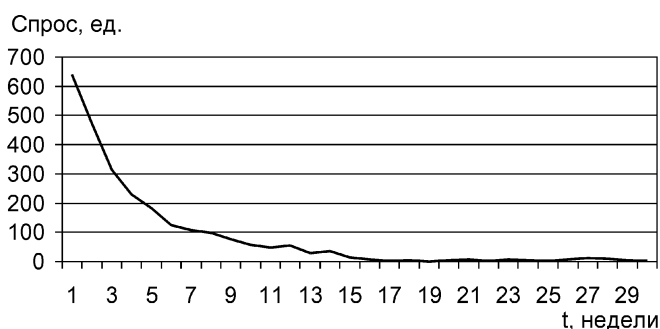


Рис. 2. График спроса на “новый объект”

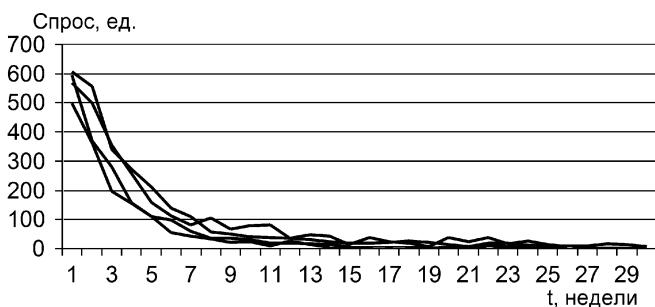


Рис. 3. Графики спроса на отобранные объекты-аналоги

каждым экспертом в качестве аналогов были подобраны несколько объектов. После ввода коэффициентов масштаба и похожести системой для каждого эксперта были сформированы прогнозы спроса на рассматриваемый объект.

Построив прогноз на основе выбранных аналогов и введенных коэффициентов, система в ряде случаев рекомендовала внести в списки некоторые изменения. Например, на рис. 3 показаны графики временных рядов объектов, отобранных одним из экспертов в качестве аналогов.

Один из объектов в списке аналогов отдалён от построенного прогноза больше, чем другие. Система порекомендовала эксперту удалить данный объект из списка аналогов. Кроме того, система предложила эксперту список объектов, временные ряды которых достаточно близки к сформированному прогнозу.

В результате выполнения описанной выше процедуры каждый из принимавших участие в процедуре прогнозирования экспертов сформировал свой собственный список аналогов и предварительный прогноз временного ряда. Данная информация, а также все необходимые комментарии были сохранены в базе данных системы.

Как сказано выше, итоговый прогноз формируется и утверждается главным экспертом компании. Как правило, подобная персона имеется во всех организациях. В разделе главного эксперта ему были представлены все прогнозы рядовых экспер-

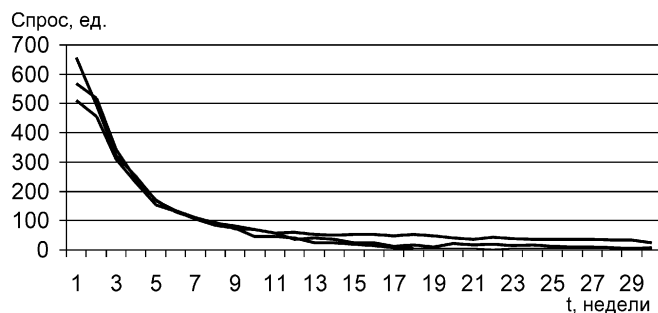


Рис. 4. Графики прогнозов, сформированных рядовыми экспертами

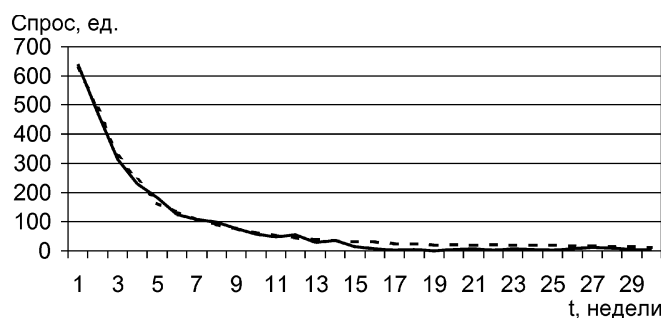


Рис. 5. Графики прогноза, сформированного главным экспертом

тов компании, а также списки аналогов, сформированные этими экспертами. На рис. 4 показаны графики сформированных экспертами прогнозов.

Главный эксперт с помощью средств раздела главного эксперта проанализировал прогнозы рядовых экспертов и внёс в них небольшие поправки. Корректировки касались, в первую очередь, коэффициентов доверия к экспертам. На рис. 4 видно, что прогнозы экспертов достаточно близки друг к другу, поэтому главный эксперт (с помощью изменения коэффициентов доверия) “пододвинул” результирующий прогнозный ряд к прогнозу одного из экспертов.

Графики итогового прогноза и фактически полученных значений временного ряда приведены на рис. 5. Штриховой линией представлены значения прогнозного ряда, сплошной – фактически полученные значения ряда. Как видно из рисунка, класс, к которому принадлежал объект прогнозирования, а также объекты-аналоги были подобраны рядовыми экспертами правильно, и прогноз оказался достаточно точным.

ЭКСПРИМ – СИСТЕМА ИМИТАЦИОННОГО МОДЕЛИРОВАНИЯ ДЛЯ ВЕРИФИКАЦИИ СИСТЕМЫ ЭКСПАМ

Перед тем, как рекомендовать системы, подобные ЭКСПАМ, к внедрению, чрезвычайно важно

убедиться в том, что они эффективны. Проблема верификации эффективности и, в частности, подтверждения достоверности выводов, формируемых с помощью экспертно-статистических систем, достаточно сложна, поскольку организовать статистически значимое экспериментальное исследование таких систем с участием высококвалифицированных экспертов, отвлекая их от основной работы, весьма непросто.

В качестве одного из альтернативных способов проверки достоверности рекомендаций, формируемых экспертно-статистическими системами, предлагаются процедуры имитационного моделирования, реализованные в форме системы ЭКСПРИМ.

В систему ЭКСПРИМ встроена 16-параметрическая модель поведения эксперта, которая позволяет имитировать различные уровни профессионализма и психологические типы экспертов (от оптимистов до пессимистов) при решении ими проблем прогнозирования спроса на вновь выпускаемую продукцию на базе применения экспертно-статистической системы поддержки процесса прогнозирования ЭКСПАМ. Более подробное описание результатов экспериментов с системами ЭКСПРИМ и ЭКСПАМ можно найти в брошюрах [3, 4].

Ниже мы ограничимся тем, что приведём несколько таблиц и графиков, построенных с помощью системы ЭКСПРИМ.

При моделировании эксперта для описания его профессионализма были введены параметры первого (Ep_1) и второго (Ep_2) рода. Параметр профессионализма первого рода характеризует способность эксперта к правильному назначению объектов-аналогов, он изменяется от 1 до 5 (от абсолютной некомпетентности до идеального выбора). Параметр профессионализма второго рода характеризует способность эксперта к точному формированию и корректировке самих прогнозов, “угадыванию” (он также изменяется от 1 до 5: от грубых промахов до идеального “угадывания” – не эксперт, а “волшебник”).

Из табл. 1 видно, что с ростом каждой из степеней профессионализма (Ep_1 , Ep_2) среднее значение ошибки прогноза уменьшается, причём основной вклад вносит увеличение значения Ep_1 , т. е. параметра, характеризующего способность эксперта правильно подбирать объекты, похожие на ОП. Меньший вклад параметра Ep_2 , т. е. способности эксперта угадывать значения прогнозируемого ряда, объясняется тем, что эксперт производит ручную корректировку значений ряда с вероятностью, не равной единице, а – стало быть – не всегда.

Из табл. 2 ясно, что для эксперта с наименьшим значением параметра Ep_1 среднеквадратическое отклонение ошибки прогноза растёт с увели-

чением значения параметра Er_2 . Причина такого роста понятна. Среднее значение ошибки прогноза велико, т. е. прогноз сильно отклонён от фактического значения ряда. Однако с ростом значения параметра Er_2 увеличивается вероятность того, что эксперт, несмотря на некорректный подбор аналогов, всё-таки более или менее точно угадает будущее значение ряда. Но так как ручная корректировка осуществляется с отличной от единицы вероятностью, то получается, что эксперт в качестве прогноза может указать как сильно удалённое от фактических данных значение (т. е. не сможет не скорректировать прогноз, выполненный на основе некорректных аналогов), так и достаточно близкое к фактическим данным значение прогноза. Это и приводит к росту среднеквадратического отклонения ошибки прогноза с увеличением значения параметра Er_2 при наименьшем значении параметра Er_1 .

С ростом значения параметра Er_1 данная зависимость становится менее выраженной, а с приближением параметра Er_2 к максимальному значению и вовсе “исчезает”. Это можно объяснить тем, что прогноз, сформированный экспертом на основе аналогов, становится достаточно близким к фактическому значению ряда, а ручная корректи-

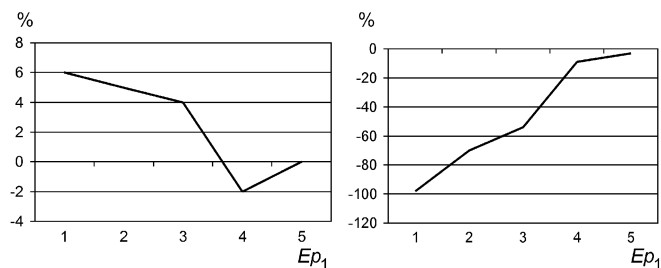


Рис. 6. Зависимость среднего значения ошибки прогноза для первой точки от значения параметра Er_1

ровка, если и осуществляется, то, как правило, на отрезке между прогнозом и фактическим значением ряда. Поскольку с ростом параметра Er_1 длина данного отрезка уменьшается, то дисперсия ошибки практически перестаёт зависеть от значения параметра Er_2 .

Приведенные результаты получены для объекта, по которому в базе данных системы имеется очень небольшое число аналогов. Для объектов, по которым в базе данных содержится значительное число аналогов, ситуация несколько иная. Дело в том, что, даже если выбор аналогов осуществляется непрофессиональным экспертом, вероятность отбора объектов из близких к ОП групп похожести достаточно велика. К примеру, на рис. 6 приведены графики для двух объектов, по которым в базе данных имелось значительное число аналогов.

Более полное описание многочисленных экспериментов с системой ЭКСПРИМ можно найти в брошюре [4]. Эти эксперименты доказывают, что предложенный подход к прогнозированию временных рядов по коротким выборкам позволяет получать адекватные (и по точности прогноза, и по числу необходимых для этого шагов) решения задач прогнозирования.

ЛИТЕРАТУРА

1. Мандель А. С. Метод аналогов в прогнозировании коротких временных рядов: экспертно-статистический подход // Автоматика и телемеханика. — 2004 (в печати).
2. Бауман Е. В., Дорофеев А. А. Классификационный анализ данных / В кн.: Избр. тр. Междунар. конф. по проблемам управления. Т. 1. — М.: СИНТЕГ, 1999.
3. Беляков А. Г., Мандель А. С. Прогнозирование временных рядов на основе метода аналогов (элементы теории экспертно-статистических систем). Препринт. — М.: Институт проблем управления, 2002.
4. Беляков А. Г., Мандель А. С. Анализ достоверности выводов, формируемых с помощью экспертно-статистических систем. Препринт. — М.: Институт проблем управления, 2002.

☎ (095) 334-89-69

E-mail: manfoon@ipu.rssi.ru



Таблица 1

Зависимость среднего значения ошибки прогноза для первой точки ряда от степеней профессионализма эксперта Er_1 и Er_2

Er_1	Er_2				
	1	2	3	4	5
1	-587	-507	-433	-340	-292
2	-420	-411	-301	-232	-188
3	-265	-267	-198	-159	-125
4	-82	-78	-38	-37	-35
5	-39	-20	-20	-19	-8

Таблица 2

Зависимость среднеквадратического отклонения ошибки прогноза для первой точки ряда от степеней профессионализма эксперта Er_1 и Er_2

Er_1	Er_2				
	1	2	3	4	5
1	88	114	167	226	296
2	161	125	175	189	249
3	169	143	142	152	182
4	86	73	73	71	62
5	55	53	45	51	47