

Abstract. Article contains a review of existing decision support systems and modeling approaches used in forecasting of business processes. It describes the problems and reasons for the creation of specialized systems for solving similar problems with different types of formalization of quality criteria. Experimental results of comparing the created DSS that show savings of working time of business analyst by 20%–30% when forecasts are made with insufficient formalization of quality criteria.

Key words: DSS, forecasting, formalization of quality criteria, econometrics.

Аномація. Стаття містить огляд існуючих систем підтримки прийняття рішень та підходів до моделювання, що використовуються при прогнозуванні бізнес-процесів. Описано проблеми та причини створення спеціалізованих систем для вирішення подібних задач при різних типах формалізації критеріїв якості. Наведено експериментальні результати порівняння розробленої СППР, що показують економію робочого часу бізнес-аналітика на 20%–30% при створенні прогнозів з недостатньою формалізацією критеріїв якості.

Ключові слова: СППР, прогнозування, формалізація критеріїв якості, економетрика.

Аннотация. Статья содержит обзор существующих систем поддержки принятия решений и подходов к моделированию, используемых при прогнозировании бизнес-процессов. Описаны проблемы и причины создания специализированных систем для решения подобных задач при различных типах формализации критериев качества. Приведены экспериментальные результаты сравнения разработанной СППР, показывающие экономию рабочего времени бизнес-аналитика на 20%–30% при создании прогнозов с недостаточной формализацией критериев качества.

Ключевые слова: СППР, прогнозирование, формализация критериев качества, эконометрика.

1. Введение

Человек всегда хотел узнать будущее, поэтому профессия предсказателя была востребована во все века. На протяжении тысячелетий ею занимались гадалки, маги, астрологи, философы, используя для предсказаний кости, руны, звезды, логику. Составление прогнозов востребовано так же, как и раньше, а, может быть, даже и больше, но теперь для этого используют математический аппарат, а предсказателей уверенно потеснили аналитики. Особенно большим спросом пользуются прогнозы в области экономики [1], поскольку они влияют на политику компаний и, как следствие, на получаемую прибыль.

Получение прогнозов экономических факторов есть одна из задач эконометрики. Существуют несколько подходов для получения таких прогнозов. Самый простой – это узнать мнение эксперта (или группы экспертов) в данной области, который, на основании собственного многолетнего опыта и анализа текущей ситуации, может сделать предсказание интересующих факторов. Данный метод широко используется, но не всегда удобен по ряду причин. Во-первых, тяжело оценить степень доверия к таким предсказаниям, поскольку они носят субъективный характер: во-вторых, эксперту зачастую легко дать качественную оценку поведения интересующего параметра, но затруднительно количественную [2]. Другой подход к получению прогноза – это создание математической модели экономического процесса [3]. Специалиста, который создает такие модели, будем называть бизнес-аналитиком (БА). При создании модели БА все равно пользуется услугами эксперта, поскольку только эксперт может определить, хотя бы приблизительно, какие факторы и как влияют на данный процесс, оценить качество построенной модели и полученных результатов. Данный подход к получению прогнозов является более предпочтительным, поскольку дает возможность получить количественные оценки прогнозируемых параметров, позволяет вычислить доверительные интервалы и другие параметры, характеризующие качество модели. При построении моделей БА стал-

живается со следующими проблемами: экономические параметры часто имеют значительную погрешность, многие факторы являются неизмеримыми или данные о них отсутствуют, во многих случаях длина временного ряда коротка, факторы, используемые в модели, разнородны, нестационарные и зависят друг от друга [1]. Все это выдвигает дополнительные требования к программным продуктам, которые использует БА.

В данной работе мы хотим сделать обзор существующих программных систем, которые могут быть использованы БА в качестве системы поддержки принятия решения (СППР) и выявить их недостатки, описать требования, которым должна удовлетворять специализированная СППР БА.

2. Постановка задачи

Необходимо составить прогноз экономического фактора на заданном интервале. Экономический фактор может быть произвольного типа, например, уровень потребления или цена заданного товара, объем производства или валовой внутренний продукт страны, уровень безработицы или инфляции и т.п. Для построения прогноза необходимо создать математическую модель соответствующего экономического процесса и из всех возможных моделей выбрать ту, которая наилучшим образом удовлетворяет требованиям, заданным бизнес-экспертом.

Исходными данными для построения модели служат:

- исторические данные прогнозируемого параметра;
- неформализованное описание экспертом связей данного фактора с другими экономическими факторами, которые могут на него влиять;
- исторические данные и прогноз экономических факторов, влияющих на прогнозируемый;
- неформализованное описание дополнительных требований к модели и прогнозу.

Результатом работы БА является:

- модель бизнес-процесса в виде уравнения, удовлетворяющая всем заданным требованиям бизнес-эксперта;
- стандартные параметры модели, позволяющие оценить ее качество (среднеквадратичная ошибка, R^2 , доверительные интервалы и т.п.);
- прогноз на заданном интервале.

Основной задачей СППР является минимизация времени, затраченного БА на получение окончательного результата.

3. Обзор существующих систем прогнозирования и анализа данных

На сегодняшний день рынок программных продуктов предлагает большое количество программных пакетов, позволяющих решать поставленную задачу. Нами были опробованы следующие программные продукты: SAS [4], Statistica 8 [5], SPSS 17 [6], Statgraphics Centurion XV [7], Eviews 5.1. plus [8], Matlab [9] и специализированная СППР TrendCaster. Одни из этих программных продуктов специализируются на анализе данных и построении прогнозов (Statistica, Statgraphics, Eviews), другие являются универсальными системами работы с данными, способными, в том числе, на построение модели для прогноза (Matlab, SAS, SPSS). Одни являются платными, другие условно плат-

ными, третьи бесплатными. Все, без исключения, содержат большое количество (десятки) типов моделей. Краткая таблица основных характеристик данных систем представлена в табл. 1.

Таблица 1. Программные продукты и критерии их оценки

Критерии оценки	Statgraphics	SPSS	Statistica	Eviews	SAS	Matlab	Trend Caster
1	2	3	4	5	6	7	8
Наличие основных методов прогнозирования	+	+	+	+	+	+	+
Возможность самостоятельного задания вида модели	+	+	+	+	+	+	+
Минимальная длина истории	7	6	2	2	2	2	2
Время, необходимое для построения прогноза, по пятибалльной шкале (1 – быстро; 5 – медленно)	2	4	4	5	5	3	1
Вывод коэффициентов уравнения	+	+	+	+	+	+	+
Удобство в/в данных и параметров модели по пятибалльной шкале (1 – удобно; 5 – неудобно)	4	4	3	5	5	3	1
Графическая визуализация прогнозных и исходных данных	+	+	+	+	+	+	+
Вывод и визуализация стандартных критериев качества модели	+	+	+	+	+	+	+

Какие из данных критериев являются основными для БА? Постараемся взглянуть на эти программные продукты его глазами.

Существующие подходы к созданию модели можно разделить на три типа.

1) К первому типу отнесем случаи, когда при создании модели используются стандартные критерии качества, такие как среднееквадратичное отклонение, корреляция и т.п., а адекватность прогноза оценивается только по доверительным интервалам. В данном случае оптимальным является использование одного из пакетов SAS, Statistica, Statgraphics, Eviews, которые позволяют строить модели и прогнозы полностью в автономном режиме. При этом от пользователя СППР не требуется иметь хорошей подготовки в области моделирования, поскольку выбор типа модели и ее параметров происходит, практически, без его участия. Хорошее знание СППР также не обязательно, поскольку большинство из них имеют интуитивно понятный интерфейс пользователя для автоматического режима работы. Время на построение одной модели исчисляется секундами, и основные затраты времени приходятся на ввод-вывод данных.

2) Ко второму типу отнесем случаи, когда критерии выбора модели не являются стандартными, но достаточно хорошо формализованы. В этом случае оптимальными пакетами являются SAS, SPSS, Matlab, поскольку они содержат внутренний язык программирования, на котором легко реализуются как модели, так и, возможно, программирование критериев их отбора. В этом случае пользователь СППР должен обладать хорошим знанием самой СППР и внутреннего языка программирования, а также должен хорошо понимать, какие модели могут быть использованы в том или ином случае, поскольку от этого напрямую зависит сложность самой программы. В этом случае

у нас достаточно много времени на подготовительную работу (программирование разных вариантов моделей и критериев их отбора). После этого время на построение одной модели, как и в предыдущем варианте, исчисляется секундами. Если необходимо построить прогнозы с одинаковыми критериями качества, то для группы процессов этот подход является оптимальным, поскольку подготовительная работа не зависит от размеров группы, а зависит только от количества критериев. При малом размере группы, ввиду относительно больших затрат времени, аналитики, как правило, пользуются другим подходом к созданию модели. Поскольку с помощью внутренних языков, как правило, достаточно легко загружать и сохранять данные в более-менее произвольном формате, то проблем с вводом-выводом обычно не возникает. При выборе СППР для решения данного типа задач аналитик будет руководствоваться в первую очередь знанием (или возможностью изучения) той или иной системы и ее доступностью (подавляющее большинство таких систем платные, и стоимость их достаточно велика).

3) К третьему типу отнесем случаи, когда критерии выбора модели плохо формализованы или же мы имеем хорошо формализованные критерии, но размер группы, для которой они определены, очень мал, что делает подход второго типа нецелесообразным. Этот случай является самым сложным для аналитика, поскольку требует от него наибольшей квалификации. При этом подходе аналитик самостоятельно определяет набор типов моделей и их параметров, из которых будет выбран оптимальный вариант. При анализе результатов моделирования и прогнозирования аналитик самостоятельно контролирует все критерии, заданные экспертом. Время на создание одной модели зависит от сложности задачи и удобства работы в СППР и обычно занимает от 1 минуты до 1 часа, в среднем 5-7 минут. В данном подходе СППР играет ключевую роль, поскольку для задач одинаковой сложности затраченное время будет зависеть от того, насколько удобно будет аналитику генерировать модели и контролировать их качество. При реализации данного подхода существуют несколько стратегий:

1. Первая заключается в том, что для всех процессов модели определяются аналитиком в ручном режиме.

2. Вторая стратегия заключается в том, что сначала аналитик строит все модели в автоматическом режиме, при этом, как правило, для части процессов модели могут быть признаны удовлетворяющими необходимым критериям, для остальных же процессов модели опять подбираются в ручном режиме.

3. Третья стратегия заключается в том, что аналитик в автоматическом режиме для каждого процесса генерирует некоторое количество моделей, удовлетворяющее части критериев, которые легко формализовать, и из построенных для процесса моделей отбирает те, которые удовлетворяют оставшимся критериям. Если же ни одна модель не удовлетворяет всем критериям, аналитик, как и в остальных стратегиях, переходит к построению модели в ручном режиме.

Вторая и третья стратегии в одних случаях позволяют ускорить процесс моделирования, в других наоборот: являются пустой тратой времени, при этом заранее нельзя сказать, какая из стратегий даст наилучший результат. Поэтому аналитик при выборе стратегии должен руководствоваться только своим опытом.

И самое главное, какие же СППР можно рекомендовать при данном подходе? Для данного подхода годятся все вышеперечисленные СППР. В каждой из них предусмотрена возможность выбора модели пользователем; существуют средства оценки ее качества и т.д., в общем все, что необходимо аналитику. В то же время, как и любая универсальная система, каждая из них недостаточно гибкая, невозможно заранее предусмотреть все необходимые средства контроля и желательные дополнения, которые так необходимы пользователю в данной ситуации. Легко подсчитать, если среднее время для создания одного прогноза, которое в среднем занимает 5–7 минут, изменится на одну минуту. Это оборачивается для аналитика выигрышем или потерей 10–20% рабочего времени, что очень существенно. Поэтому для данного подхода так часто используются специализированные СППР, которые уступают общеизвестным программным пакетам и по количеству типов моделей, и развитости пользовательского интерфейса, и количеству форматов ввода-вывода, но в то же время приспособленные для решения определенного типа задач и позволяющие сэкономить рабочее время БА.

4. Выбор оптимальной модели

Так что же это за такие «плохо формализуемые» критерии, из-за которых нужно создавать специализированные системы и почему нельзя ограничиться стандартными? В данном разделе мы постараемся более подробно коснуться этого вопроса, поскольку он так существенно влияет на выбор СППР.

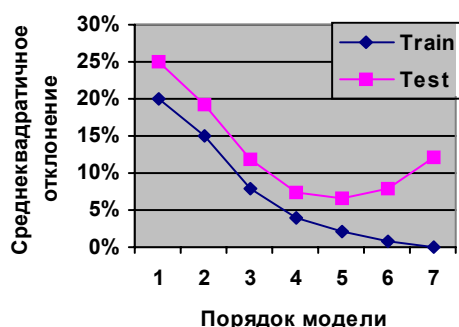


Рис. 1. Зависимость качества работы от сложности модели на обучающей и тестовой выборках

Существует великое множество методов построения модели для временных рядов: полиномиальная регрессия, векторная регрессия, авторегрессия, ARIMA, МГУА, нейронные сети и многие другие [3, 10]. При помощи каждого метода, увеличивая сложность модели, можно построить модель процесса с нулевым среднеквадратичным отклонением на истории. Но будет ли такая модель адекватно отражать процесс? Как хорошо показано в работах [10], чрезмерное увеличение сложности модели зачастую ведет к ухудшению качества самой модели. На рис. 1 показан схематический график зависимости среднеквадратичного отклонения на участках временного ряда: Train – участок, по которому строилась модель; Test – участок, который не участвовал при построении модели, на данном участке проводилось ее тестирование.

Порядок модели отражает ее сложность, например, для регрессий – это количество членов уравнения. Видно, что до определенного порядка модели среднеквадратичное отклонение падает на обоих участках, но после некоторого порогового значения среднеквадратичное отклонение на участке Train продолжает падать вплоть до нуля, в то же время на участке Test оно начинает расти. Это явление называют потерей моделью обобщающих свойств в нейронных сетях; оно также известно под названием «оверфитинг».

Для определения оптимального порядка модели, если данных достаточно, можно, как показано выше, разбить временной ряд на два участка и определить точку оптимума экспериментально. К сожалению, в реальных задачах данных обычно не хватает, поэтому уменьшение размеров временного ряда, по которому происходит построение модели, нецелесообразно. Исходя из этого, для оценки адекватности модели реальному процессу эксперт может сформулировать ряд дополнительных требований, которые, в силу человеческой психологии, обычно носят качественный характер и не являются строго формализуемыми [2]. Например, если эксперт утверждает, что рост цены на товар будет несколько опережать инфляцию или увеличение безработицы скажется умеренно отрицательно на потреблении заданного товара, БА может это учесть при выборе модели либо используя свой опыт, либо предоставив эксперту на выбор несколько моделей, удовлетворяющих данным требованиям. Главное, что, имея даже такие неточные критерии качества, БА может отсеять большинство заведомо неадекватных моделей и тем самым сократить время получения окончательного варианта.

5. Основные требования специализированной СППР

В данном разделе мы попытаемся сформулировать основные требования к специализированной СППР, которые бы максимально увеличивали эффективность работы БА.

1) Алгоритмы моделирования. Мы не будем перечислять поименно названия всех необходимых алгоритмов, поскольку определить, какие из них являются лучшими без учета задачи, нельзя. Однако можно сформулировать требования к группам алгоритмов, которые обязательно должны присутствовать в СППР:

- регрессионные модели, позволяющие учитывать влияние различных факторов;
- модели, учитывающие сезонность временного ряда;
- модели, которые могут быть настроены по коротким временным последовательностям,

поскольку в ряде случаев исходных данных может быть мало, а многие модели требуют для настройки параметров большое количество точек;

- методы, позволяющие работать с большим количеством данных (нейронные сети, МГУА).

2) Моделирование процесса:

- автоматический выбор оптимальной модели и ее параметров;
- удобный выбор модели и ее параметров в ручном режиме.

3) Инструменты контроля качества модели и прогноза:

- визуализация исторических данных и прогноза;
- визуализация стандартных критериев оценки качества модели (среднеквадратичное отклонение, корреляция, доверительные интервалы и т.д.);

• методы контроля сезонности, среднего роста и других параметров, которые удобны аналитику при решении его задачи;

- удобное добавление специализированных критериев качества.

4) Поскольку в исходных данных зачастую присутствует шум в виде выбросов (артефактов), в СППР должна быть предусмотрена возможность контроля и фильтрации артефактов.

5) Ввод-вывод данных:

- СППР должна иметь набор входных форматов данных либо один формат входных данных, в который легко конвертировать другие форматы при помощи стандартных средств работы с данными;

- вывод уравнения модели и стандартных критериев качества модели;
- результаты прогноза на заданном интервале.

6. Экспериментальные результаты

По данному принципу нами была разработана специализированная СППР “TrendCaster” (рис. 2), которая разрабатывалась и опробовалась при прогнозировании экономических факторов, таких, как макроэкономические показатели и объемы продаж товаров массового потребления, демографических и социологических параметров для ряда стран. Испытание данной СППР проходили на протяжении более 3 лет. Общее количество время использования данной СППР аналитиками нашей группы составило около 15 тысяч часов. Данные работы проводились по заказу компании “4iCG CONSULTING GROUP INC.” Эксперты, оценивающие качество прогноза, также предоставлялись данной компанией.

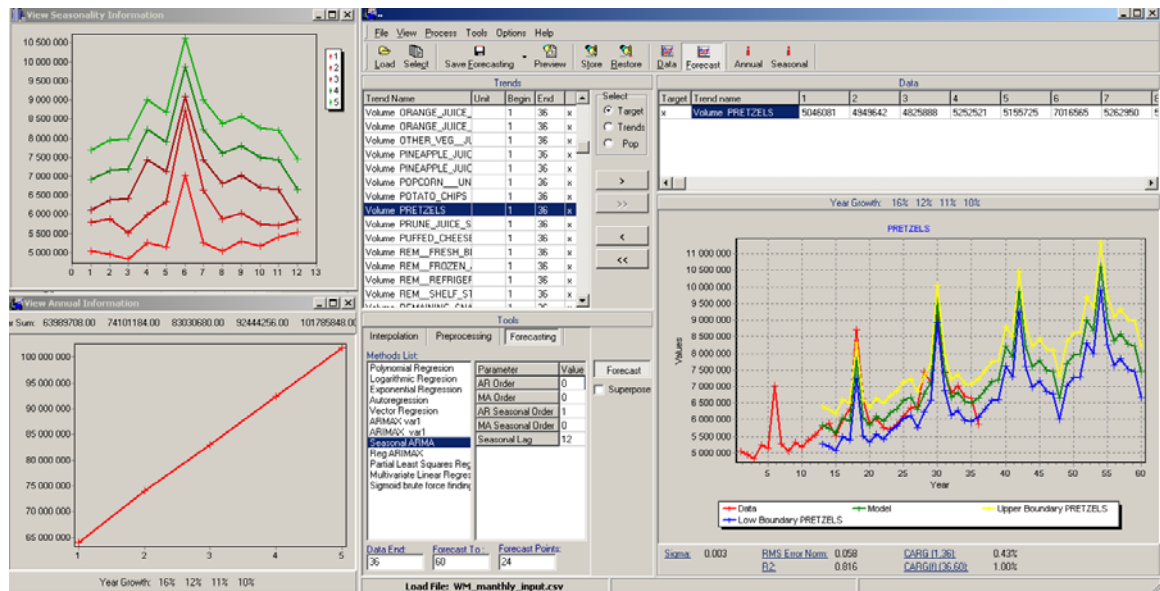


Рис. 2. Графический интерфейс пользователя СППР “TrendCaster”

Основной поток задач данного заказчика составляют прогнозы с плохо формализованными критериями качества.

Приведем один из примеров решения типичной задачи с использованием SAS и СППР “TrendCaster”. Было необходимо составить 200 прогнозов для категорий, отражающих уровень продаж товаров массового потребления. От экспертов были получены общие рекомендации по оценке качества прогнозов для товаров данной группы.

Технология получения прогнозов с использованием SAS проходила следующие этапы:

- составление прогнозов для всей группы в автоматическом режиме (10 мин.);

- проверка БА каждой модели на соответствие рекомендациям эксперта (600 мин.). Для данной задачи 120 из 200 (60%) категорий были признаны БА как удовлетворяющие требованиям эксперта;

- подбор БА моделей для 80 оставшихся категорий (1200 мин.);
- итого, общий процесс получения результата занял 1810 мин. (~30 ч).

При построении прогноза с использованием СППР "TrendCaster" БА самостоятельно для каждой категории выбирает тип и параметры модели. Поскольку в СППР "TrendCaster" реализован удобный и быстрый переход от одного типа модели к другому, а в рабочем окне программы отражается более полный набор различных факторов, позволяющих контролировать соответствие модели рекомендациям эксперта, БА затрачивает меньше времени на выбор оптимальной модели. Для данной работы общее время, потраченное на получение 200 прогнозов, составило 1100 мин. (~18,5 ч), что на 39% меньше, чем потребовалось БА с использованием SAS. Конечно, величина выигрыша очень зависит от решаемых задач, однако по результатам нашей работы мы можем утверждать, что при построении моделей, для которых критерии качества плохо формализованы, СППР "TrendCaster" позволяет экономить 20%–30% рабочего времени БА.

7. Выводы

В данной работе мы сделали обзор существующих стандартных программных продуктов, которые используются для построения эконометрических прогнозов. Обосновали причины, по которым создаются аналогичные специализированные системы. По итогам испытания одной из таких систем (СППР «TrendCaster») был оценен выигрыш в экономии рабочего времени БА, который составил 20%–30% по сравнению с использованием стандартных средств, при построении прогнозов с плохо формализованными критериями качества.

СПИСОК ЛИТЕРАТУРЫ

1. Хэндри Д. Эконометрика: алхимия или наука? / Д. Хэндри // Эковест. – 2003. – № 2. – С. 172 – 196.
2. Орлов А.И. Устойчивость в социально-экономических моделях / Орлов А.И. – М.: Наука, 1979. – 296 с.
3. Цыплаков А. Введение в прогнозирование в классических моделях временных рядов (рус.) / А. Цыплаков // Квантиль. – 2006. – № 1. – С. 3 – 19.
4. SAS, Statistical Analysis System, Student Supply Store / Barr, J. Anthony, Goodnight, H. James H. [et al.]. – North Carolina State University, 1971. – OCLC 5728643.
5. Joaquim Sá Applied Statistics Using Spss, Statistica, Matlab and R. – Berlin: Springer, 2007. – 506 p.
6. <http://www.statgraphics.com>.
7. <http://www.eviews.com>.
8. Бююль А. SPSS: Искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей / А. Бююль, П. Цёфель; пер. с нем. – СПб.: ДиаСофтЮП, 2005. – 608 с.
9. Дьяконов В.П. Справочник по применению системы PC MATLAB / Дьяконов В.П. – М.: Физматлит, 1993. – С. 112.
10. Ивахненко А.Г. Самоорганизация прогнозирующих моделей / А.Г. Ивахненко, И.А. Мюллер. – К.: Техника, 1985. – 350 с.

Стаття надійшла до редакції 18.05.2010