

УДК 681.31(031)

МЕТОДЫ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА И ПРОГНОЗИРОВАНИЯ ДАННЫХ СТАЦИОНАРНОЙ РОЗНИЧНОЙ ТОРГОВОЙ СЕТИ

Ребенок И.И.¹, Малыхина М.П.¹

¹ФГБОУ ВПО «Кубанский государственный технологический университет», Краснодар, Россия (350072, г. Краснодар, ул. Московская, д. 2), e-mail: malpema@mail.ru, xvilinya@gmail.com

Дано обоснование актуальности задачи интеллектуального анализа и прогнозирования данных в розничной торговле. Исследованы методы интеллектуального анализа и прогнозирования в сфере розничной торговли на основе имеющейся базы данных продаж за некоторый промежуток времени для объективной оценки будущих закупок и продаж. На основе спрогнозированных данных и реальных продаж рассчитана погрешность результатов исследования. Рассмотрены теоретические и практические аспекты интеллектуального анализа данных. Для анализа использовалась база данных реального объекта. Рассмотрен анализ временных рядов. Рассмотренные методы интеллектуального анализа могут существенно облегчить анализ данных и увеличить прибыль предприятия. В заключение приведены выводы по использованию регрессионного анализа и анализа временных рядов и оценена эффективность каждого из методов.

Ключевые слова: интеллектуальный анализ, регрессионный анализ, временные ряды, розничная торговля, хранилище данных.

DATA MINING METHODS AND FORECASTING DATA IN FIXED RETAIL CHAIN

Rebenok I.I.¹, Malykhina M.P.¹

¹Kuban State Technological University, Krasnodar, Russia (350072, Krasnodar, Moskovskaya street, 2), e-mail: malpema@mail.ru, xvilinya@gmail.com

The actuality of the task in research of data mining methods and forecasting data in fixed retail trade is justified. Data mining methods and forecasting in the retail trade based on existing sales database for a period of time for objective assessment of future purchases and sales are researched. Based on the predicted and actual sales data calculated error in the results of research. Theoretical and practical aspects of data mining methods are considered. For the analysis was used the database from real object. Time series analysis is considered. Considered data mining methods can greatly facilitate the analysis of data and increase the profits of the enterprise. In conclusion presents the findings on the use of regression analysis and time series analysis and assessed the effectiveness of each method.

Keywords: data mining, regression analysis, time series, retail chain, data warehouse.

Введение

Розничная торговая сеть - это совокупность предприятий розничной торговли и других торговых единиц, размещенных на определенной территории в целях продажи товаров и обслуживания покупателей или находящихся под общим управлением.

Изучение розничной торговли с помощью основных классификационных признаков позволяет получить информацию о ее качестве, относительной величине и важности, о воздействии внешних факторов, а также определить основные виды розничной торговой сети (рисунок 1).



Рисунок 1 - Виды розничной сети в зависимости от условий продажи

В данной статье рассмотрена стационарная розничная торговая сеть. Эта сеть включает специально оборудованные здания (магазины), осуществляющие куплю-продажу товаров и оказание услуг покупателям для их личного, семейного и домашнего использования. Каждый магазин включает в себя автоматизированное рабочее место (АРМ), которое состоит из технических (компьютер, принтер и прочее) и программных средств (база данных и программная оболочка), позволяющих упростить и автоматизировать повседневные процессы ведения документации и продаж.

Магазин в течение дня обеспечивает бесперебойную продажу продуктов массового потребления, а перед закрытием синхронизирует данные с головной компанией.

В главном офисе происходит анализ продаж каждого объекта, на основе которого прогнозируется объем дальнейших закупок и продаж с учетом внешних факторов, таких как: социальный уровень жизни в районе нахождения магазина, возможность проведения акций, количество потребителей в часы пик и т.д.

Цели исследования: выявление наиболее оптимального метода анализа и прогнозирования продаж за некоторый промежуток времени для объективной оценки будущих закупок и продаж. В качестве рассматриваемых интеллектуальных методов использовались наиболее частые в плане применения: анализ временных рядов и регрессионный анализ.

Методы и материалы исследования: для достижения целей исследования были использованы программный комплекс Statistica 10 и среда SQL Server Business Intelligence Development Studio. В качестве исходных данных использована БД продаж реального объекта. Для более качественного анализа был спроектировано хранилище данных [5] (рисунок 2).

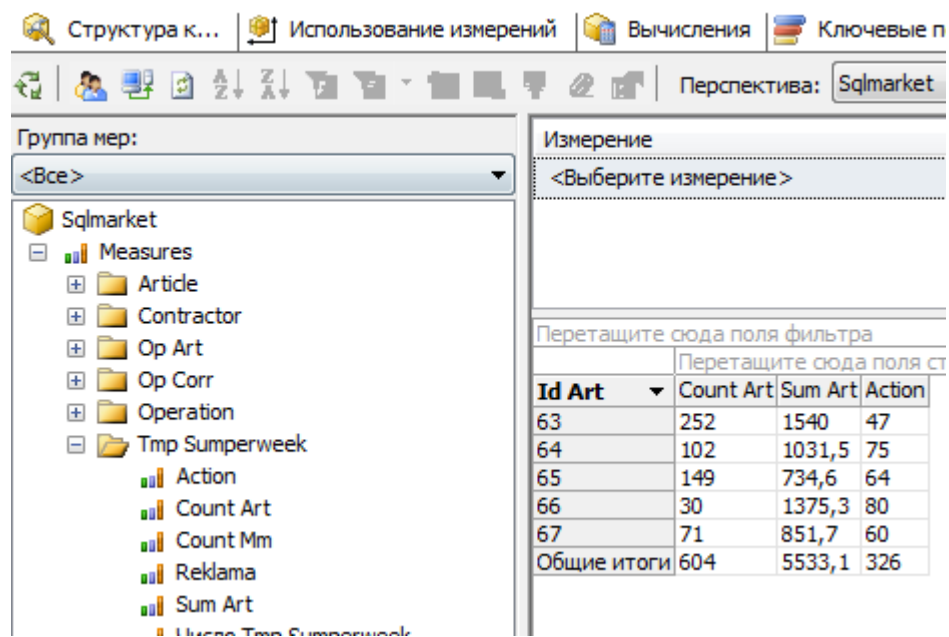


Рисунок 2 – Хранилище, представленное в виде куба данных.

Краткая характеристика рассматриваемых методов анализа и прогнозирования

1. Регрессионный анализ, задачей которого является определение аналитического выражения, приближенно описывающего зависимость одного фактора от другого за некоторый промежуток времени [2]. Форма связи между этими факторами в регрессионном анализе получила название уравнения регрессии.

$$Y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n + k \quad (1)$$

Регрессионную модель можно представить в виде графика, который называется диаграммой рассеивания (рисунок 3).

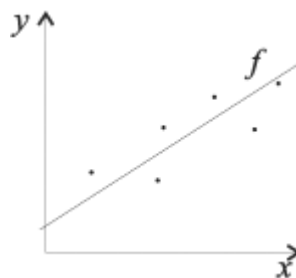


Рисунок 3 – Диаграмма рассеивания в регрессионном анализе (реальные факты – точки, регрессионная зависимость – линия).

2. Анализ временных рядов - совокупность математико-статистических методов анализа, предназначенных для выявления структуры временных рядов и для их прогнозирования [1]. Сюда относятся, в частности, методы регрессионного анализа. Выявление структуры временного ряда необходимо для того, чтобы построить математическую модель того процесса, который является источником анализируемого временного ряда (рисунок 4). Прогноз будущих значений временного ряда используется для эффективного принятия решений.

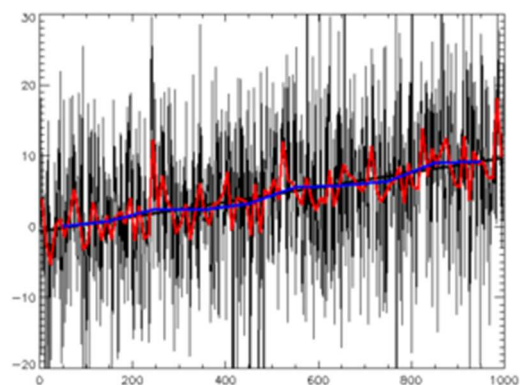


Рисунок 4 – Пример временного ряда.

Результаты исследования: регрессионный анализ реализован с помощью продукта Statistica 10. В качестве входных данных использовалась база данных продаж реального объекта, которая содержала количество проданного товара, цену, размер скидки (акция) и количество средств, использованных для рекламы продукции (рисунок 5).

| | 1 | 2 | 3 | 4 |
|----|-----------|---------|--------|---------|
| | count_art | sum_art | action | reklama |
| 1 | 30 | 280 | 2 | 3000 |
| 2 | 35 | 270 | 4 | 4000 |
| 3 | 36 | 269 | 8 | 5000 |
| 4 | 39 | 261 | 9 | 6000 |
| 5 | 40 | 260 | 11 | 7000 |
| 6 | 72 | 200 | 13 | 8000 |
| 7 | 78 | 190 | 15 | 9000 |
| 8 | 67 | 210 | 17 | 10000 |
| 9 | 56 | 230 | 18 | 11000 |
| 10 | 45 | 245 | 19 | 12000 |
| 11 | 48 | 240 | 20 | 13000 |
| 12 | 52 | 235 | 24 | 14000 |
| 13 | 55 | 225 | 27 | 15000 |
| 14 | 61 | 220 | 28 | 16000 |
| 15 | 64 | 217 | 29 | 17000 |
| 16 | 67 | 215 | 30 | 18000 |
| 17 | 70 | 210 | 30 | 19000 |

Рисунок 5 – Входные данные для регрессионного анализа (1 – количество проданного товара, 2 – цена, 3 – размер скидки (акция), 4 – средства на рекламу).

Зависимой переменной выступает количество проданного товара, остальные 3 переменные – независимые, оказывающие непосредственное влияние на значение зависимой переменной.

В результате была получена модель многофакторного регрессионного анализа, представленная графически в виде диаграммы рассеивания (рисунок 6). Точками на диаграмме рассеивания обозначены реальные продажи из базы данных продаж, проходящая линия – линия регрессии, которая позволяет увидеть общую тенденцию изменения объема продаж в зависимости от ценовой политики.

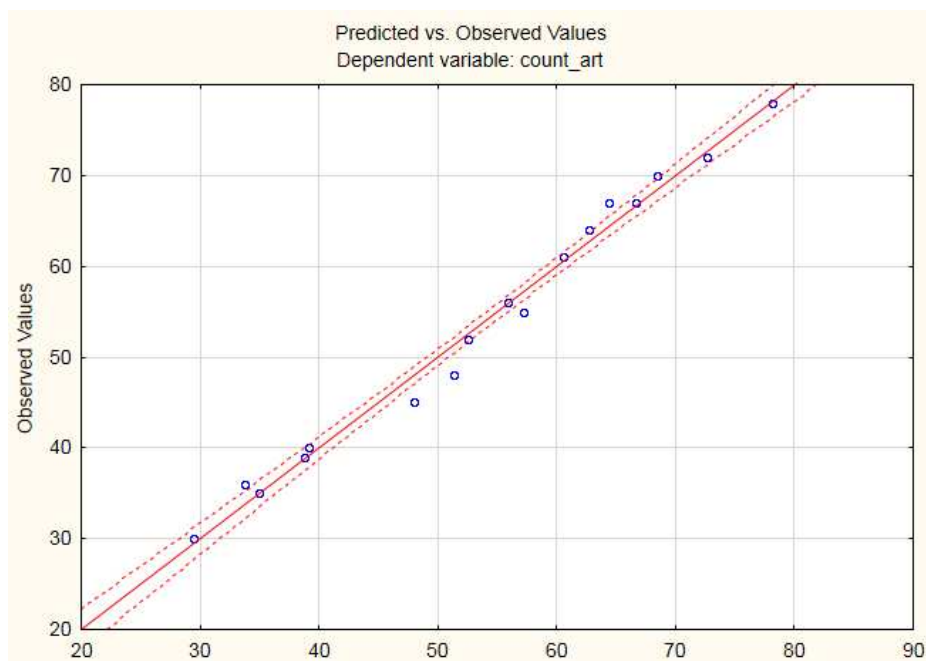


Рисунок 6 – Диаграмма рассеивания регрессионного анализа (точки на графике – реальные продажи, линия – прогнозируемые продажи).

Стоит заметить, что база реального объекта содержала данные за 2 месяца. Через две недели были запрошены данные повторно. Простое сравнение результатов показало, что прогнозирование на 1 день на основе продаж за 2 месяца с помощью регрессионного анализа дало достаточно точный прогноз, погрешность которого в сравнении с реальными продажами составила 8,6% (21 единица алкогольной продукции по прогнозу и 23 единицы по факту). При более долгосрочном прогнозировании (14 дней) погрешность составила 13,9% (271 единица алкогольной продукции по прогнозу и 315 единиц по факту).

В качестве входных параметров для анализа временных рядов использовалась та же самая база реального объекта, на основе которой создано хранилище данных. В среде SQL Server Business Intelligence Development Studio существует встроенный алгоритм для реализации анализа временных рядов, что определенно упрощает анализ любых данных [1; 3]. Входными параметрами обучающей модели анализа временных рядов выступали количество проданного товара, цена, размер скидки (акция) и количество средств, использованных на рекламу продукции. Прогнозируемыми компонентами были: количество проданного товара и цена. Результат анализа представлен графически на рисунке 7. Пунктирными линиями представлен прогноз.

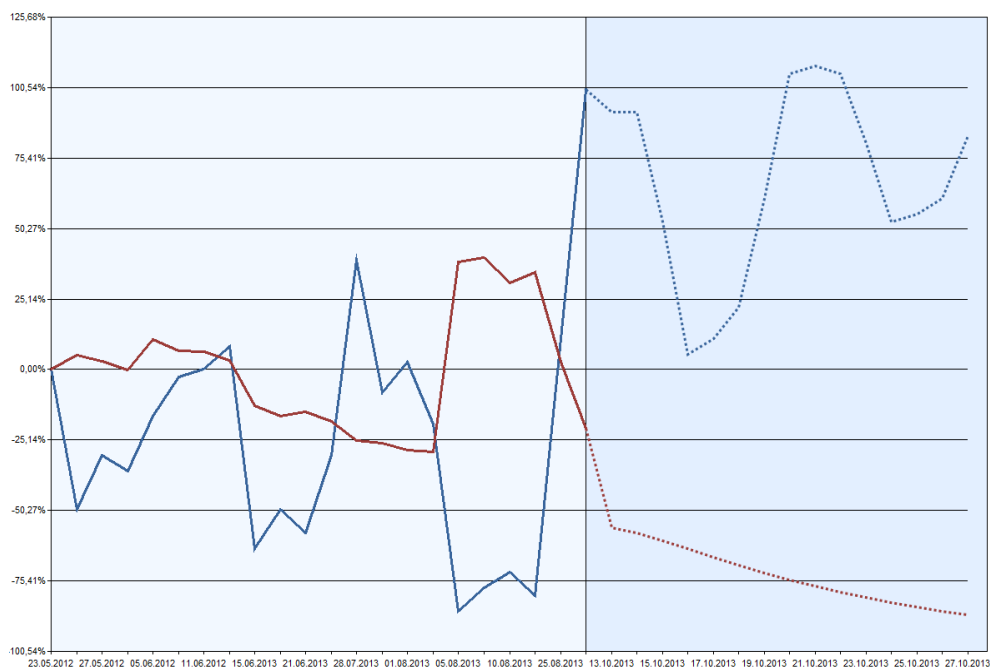


Рисунок 7 – График анализа временных рядов (красная линия – цена, синяя линия – количество проданного товара).

Исходные данные для анализа временных рядов представляли собой также продажи за 2 месяца. При прогнозировании на 1 день итоговая погрешность составила 4,3% (24 единицы алкогольной продукции по прогнозу и 23 единицы по факту). При более долгосрочном прогнозировании (14 дней) погрешность составила 2,5% (307 единиц алкогольной продукции по прогнозу и 315 единиц по факту).

Заключение

Временной ряд существенно отличается от простой выборки данных, так как при анализе учитывается взаимосвязь измерений со временем, а не только статистическое разнообразие и статистические характеристики выборки. Один из его ключевых плюсов заключается в том, что можно на основе достаточно небольшого количества данных сформировать длительный прогноз, который покажет «тренд», зависимость и влияние одних показателей на другие.

С помощью регрессионного анализа можно получить достаточно качественный прогноз. При использовании многофакторного регрессионного анализа весьма существенно увеличивалась точность прогноза по сравнению с однофакторным. Это и понятно: больше факторов для анализа – выше точность прогноза. Правда, для более качественного прогноза требуется и больший объем данных для анализа. Это справедливо для всех методов интеллектуального анализа.

Как показывают результаты проведенного анализа, временные ряды на практике эффективнее регрессионного анализа. Особенно это касается долгосрочного прогнозирования, где регрессионная модель показала себя хуже: погрешность была весьма

ощутима. К тому же анализ временных рядов позволяет одновременно обучать и прогнозировать несколько компонент (в нашем примере 2: цена и количество), что позволяет увидеть взаимное изменение и вовремя откорректировать ценовую политику.

Список литературы

1. Барсегян А.А., Куприянов М.С., Холод И.И., Тесс М.Д., Елизаров С.И. Data Mining и перегрузка информацией : вступительная статья // Анализ данных и процессов. - 3-е изд. перераб. и доп. - СПб. : БХВ-Петербург, 2009. - С. 13.
2. Дюк, Флегонтов А.В., Фомина И.К. Применение технологий интеллектуального анализа данных в естественно-научных, технических и гуманитарных областях.
3. Дюк В.А., Самойленко А. Data Mining : учебный курс (+CD). — СПб. : Питер, 2001. — 368 с.
4. Кречетов Н., Иванов П. Продукты для интеллектуального анализа данных // ComputerWeek-Москва. - 1997. - № 14-15. - С. 32-39.
5. Малыхина М.П. Базы данных: основы, проектирование, использование. Рекомендуется УМО по специальности 220400 «Программное обеспечение вычислительной техники и автоматизированных систем» для межвузовского использования. - СПб., 2004.
6. Малыхина М.П., Частикова В.А., Власов К.А. Исследование эффективности работы модифицированного генетического алгоритма в задачах комбинаторики // Современные проблемы науки и образования. - 2013. - № 3. - С. 32.
7. Afifi A., Clark V. and S. May (2003). Computer-Aided Multivariate Analysis. - 4th ed. - New York: CRC Press. - ISBN 1-58488-308-1.

Рецензенты:

Ключко В.И., д.т.н., профессор кафедры информационных систем и программирования, ФГБОУ ВПО «Кубанский государственный технологический университет», г. Краснодар.

Видовский Л.А., д.т.н., профессор, зав. кафедрой информационных систем и программирования, ФГБОУ ВПО «Кубанский государственный технологический университет», г. Краснодар.