

ВИКОРИСТАННЯ РЕГРЕСІЙНОГО АНАЛІЗУ ДЛЯ ПРОГНОЗУВАННЯ КІЛЬКОСТІ ВІДВІДУВАНЬ СУПЕРМАРКЕТІВ

Савчук Тамара, Бистрик Максим

Вінницький національний технічний університет

Анотація

Розглянуто доцільність використання регресійного аналізу з метою прогнозування кількості відвідувань супермаркетів. Знайдені закономірності допоможуть менеджерам приймати ефективні рішення при формуванні подальшої маркетингової політики організації.

Abstract

Considered expediency of using a regression analysis to predict the count of visits to the supermarkets. Found dependences, will help managers to accept effective decisions during formation the future company's marketing policy.

Вступ

Для ведення успішної торгівлі та підвищення доходів великим підприємствам важливо здійснювати прогнозування своїх прибутків. Найбільш вдалим шляхом вирішення даної задачі є прогнозування відвідувань та покупок клієнтів на основі попередньої статистики. При цьому за наявності одної чи кількох незалежних змінних X та залежної змінної Y , доцільно використовувати регресійний аналіз.

Не дивлячись на те, що автоматизовані системи управління на основі регресійного аналізу активно досліджуються, поки що питання, пов'язані з розробкою методів, моделей і алгоритмів виявлення експертних знань, класифікації ситуацій, формулювання управлінських рішень на основі статистики про відвідування супермаркетів є недостатньо вирішеними.

Зважаючи на це, для підвищення ефективності визначення стратегії маркетингової діяльності актуальною є задача розробки відповідних інтелектуальних засобів прогнозування, що базуються на методах регресійного аналізу.

При використанні лінійної регресії взаємозв'язок між даними моделюється за допомогою лінійних функцій, а невідомі параметри моделі оцінюються за вхідними даними. Подібно до інших методів регресійного аналізу лінійна регресія повертає розподіл умовної імовірності Y в залежності від X , а не розподіл спільної імовірності y та X , що стосується області мультіваріативного аналізу.

Загальна лінійна регресійна модель матиме вигляд:

$$y = \beta_0 + \beta_1 \chi_1 + \dots + \beta_k \chi_k + u \quad (1)$$

де y – залежна змінна, $(\chi_1, \chi_2, \dots, \chi_k)$ – незалежні змінні, u – випадкова похибка, розподіл якої в загальному випадку залежить від незалежних змінних але математичне сподівання якої рівне нулю.

У випадку прогнозування відвідувань супермаркетів χ_k – це вхідний незалежний вектор, який характеризує певний день тижня, а y – залежна змінна, яка відображає суму витрат покупця у цей день [1].

Інтелектуальні інформаційні технології

Відповідно згідно з цією моделлю математичне очікування залежної змінної є лінійною функцією незалежних змінних:

$$E(y) = \beta_0 + \beta_1 \chi_1 + \dots + \beta_k \chi_k + u \quad (2)$$

Вектор параметрів $(\beta_0, \beta_1 \dots \beta_k)$ – є невідомими і задача лінійної регресії полягає у оцінці цих параметрів на основі деяких експериментальних значень y і $(\chi_1, \chi_1 \dots \chi_k)$. Тобто для деяких n експериментів є відомі значення $\{y_i, \chi_{i1} \dots \chi_{ip}\}$ незалежних змінних і відповідне їм значення залежної змінної.

Згідно з визначенням моделі для кожного експериментального випадку залежність між змінними визначається формулами:

$$y_i = \beta_0 + \beta_1 \chi_{i1} + \dots + \beta_k \chi_{ki} + u \quad (3)$$

На основі цих даних потрібно оцінити значення параметрів а також розподіл випадкової величини. Зважаючи на характеристики досліджуваних змінних, можуть додаватися різні додаткові специфікації моделі і застосовуватися різні методи оцінки параметрів [2]. Серед найпоширеніших специфікацій лінійних моделей виділяють класичну модель лінійної регресії та узагальнену модель лінійної регресії. У випадку прогнозування відвідувань супермаркету краще використати класичний підхід, оскільки вхідні дані не містять великої кількості незалежних векторів

Отже, запропонована математична модель є трендовою моделлю залежності суми витрат клієнта від дня тижня, в який здійснюється відвідування супермаркету.

Розглянемо приклад поведінки покупця протягом певного періоду. На рис. 1 показана статистика відвідувань та витрат одного з клієнтів. Вона відома до 27 січня (27 січня він витратив 120 гривень, 24-го – 70, 20-го – 30 і т. д.).

Січень 18	Січень 19	Січень 20	Січень 21	Січень 22	Січень 23	Січень 24	Січень 25	Січень 26	Січень 27	Січень 28	Січень 29
	45	30				70			120	?	?

Рис. 1. Статистика покупок одного клієнта

Пн	Вт	Ср	Чт	Пт	Сб	Нд
		150	80			
100					20	
		45	30			
70			120	?	?	?

Рис. 2. Матриця тижнів

Беручи до уваги дані представлені на рис. 2, можна зробити висновок, що в деякі дні тижня покупець відвідує супермаркет частіше, ніж в інші. Залежно від того, скільки грошей клієнт витрачав під час минулих візитів у певний день, можна визначити середнє значення доходу від цього покупця при наступних візитах у ці самі дні тижня на основі вищеописаних методів пошуку оптимального рішення [3].

Таким чином, запропонований підхід дозволить ефективно спрогнозувати день, коли покупець наступного разу відвідає супермаркет та кількість грошей, яку він при цьому витратить. Використання регресійного аналізу є оптимальним рішенням для реалізації даного підходу за рахунок підвищення ефективності пошуку правильного рішення, а також дозволить отримати результат заданої точності.

Список використаних джерел:

1. Т. О. Савчук, В. В. Колодний, А. В. Козачук. Система моделювання неперервних випадкових величин з графічно заданими особливостями «GenGraph»: Міжнародний науково-технічний журнал Інформаційні технології та комп'ютерна інженерія – м. Вінниця, 2011 – №2 (21). – С.49-55.

2. Колесницький О. К. Принципи побудови архітектури спайкових нейрокомп'ютерів / О. К. Колесницький // Вісник Вінницького політехнічного інституту. – Вінниця: УНІВЕРСУМ-Вінниця. – 2014. – №4 (115), С.70-78. [Електронний ресурс]. Режим доступу - <http://visnyk.vntu.edu.ua/article/view/3697/5416>.

3. Введение в анализ данных [Електронний ресурс]. – Режим доступу: <http://alexanderdyakonov.narod.ru/lpotdyakonov.pdf>.