

УДК 662.642: 621.926.7

В.В.Меркушев (5 курс, каф. МПУ), Л.М.Яковис, к.т.н., доц.

ВЕРОЯТНОСТНО-СТАТИСТИЧЕСКИЙ АНАЛИЗ ПОСЕЩАЕМОСТИ WEB-САЙТОВ

В наши дни происходит рождение новой сферы бизнеса – Интернет коммерции. С каждым днем появляются новые web-сайты, предоставляющие самые разнообразные услуги, от развлечений и общения до заказа товаров на дом и биржевых операций. Для владельцев web-сайтов важен рост посещаемости их ресурса. С этой целью проводятся рекламные мероприятия, которые требуют оценки своей эффективности. Существующие коммерческие пакеты программ типа Rambler или SpyLOG позволяют получать различного рода статистические сводки посещаемости web-сайтов, однако не содержат средств для анализа тенденций посещаемости.

В данной работе рассматриваются возможности применения вероятностно-статистических методов для решения двух наиболее важных задач, возникающих при анализе посещаемости web-сайта. Первая из них – анализ тенденции посещаемости. Такой анализ дает представление о необходимости проведения рекламных мероприятий и об их эффективности после проведения. Другая, не менее важная задача, – определение времени изменения тенденции для своевременной реакции на это изменение.

Показатель посещаемости web-сайта (число посетителей в день) является случайным процессом с такими характерными свойствами, как тренд и сезонность. Тренд позволяет сделать выводы о тенденции изменения посещаемости, но наличие сезонности затрудняет выделение тренда. Для устранения влияния сезонности на выявление тренда разработан алгоритм, позволяющий рассчитать коэффициенты сезонности, определить их значимость и при наличии значимой сезонности нужным образом изменить значения показателя посещаемости.

Для проведения исследований использовались реальные статистические данные ежедневной Интернет-газеты Лента.Ру (www.lenta.ru), показатель посещаемости которой особо подвержен сезонности. Аудитория данного web-сайта весьма специфична, и посещаемость в выходные дни в 1,5 раза меньше посещаемости по рабочим дням (рис. 1). Результаты работы алгоритма, устраняющего влияние сезонности, приведены на рис. 2.

Для выделения трендов временных рядов показателя посещаемости web-сайта предложено использовать полиномиальные регрессионные модели, которые строятся по методу наименьших квадратов [1]. Наиболее простой моделью данного класса является линейная зависимость числа ежедневных посещений сайта от времени $\hat{y} = b_0 + b_1x$, где b_0, b_1 – параметры регрессии (рис. 2). Сравнивая модели на двух интересующих временных интервалах, можно сделать выводы о наличии или отсутствии значимых изменений среднего уровня и средней скорости дрейфа временного ряда. На основе такого анализа делаются выводы о характере тенденции показателя посещаемости.

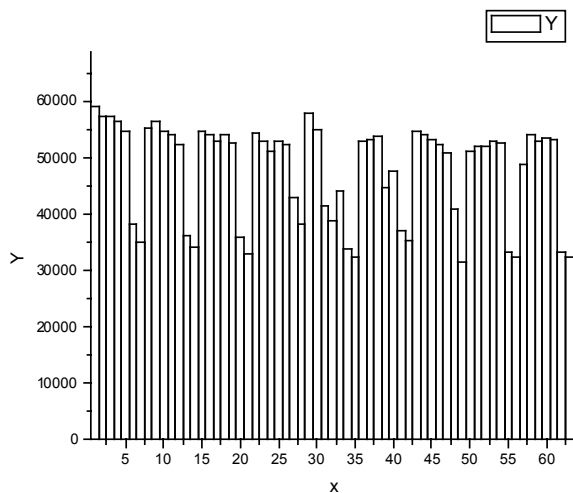


Рис. 1

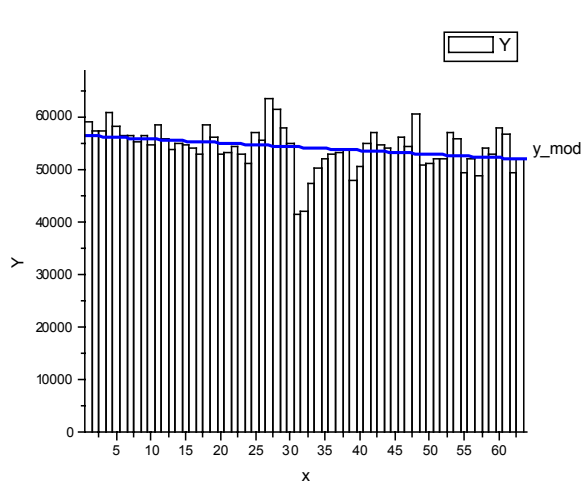


Рис. 2

Для обнаружения времени изменения тенденции посещаемости, обычно выражающейся в текущем изменении среднего уровня показателя посещаемости, предложено использовать метод экспоненциального сглаживания [2]. Сглаживание осуществляется по рекуррентной формуле $S_t = \alpha x_t + (1 - \alpha)S_{t-1}$, где S_t – экспоненциальная средняя. Экспоненциальное сглаживание можно представить как фильтр, на вход которого подается исходный ряд, а на выходе формируются текущие значения экспоненциальной средней. Значения экспоненциальной средней для показателя посещаемости web-сайта “Омен” (www.omen.ru) представлены на рис. 3.

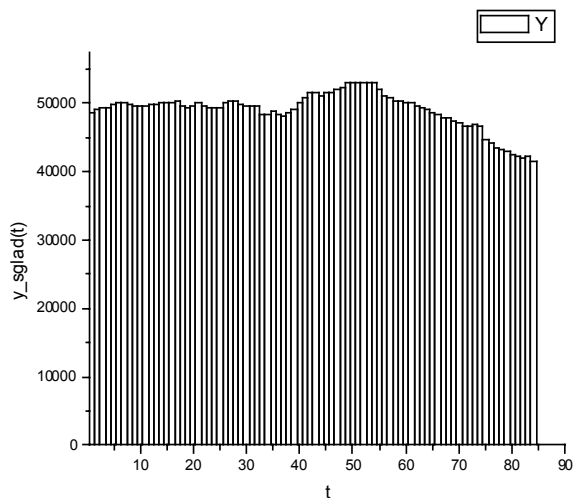


Рис. 3

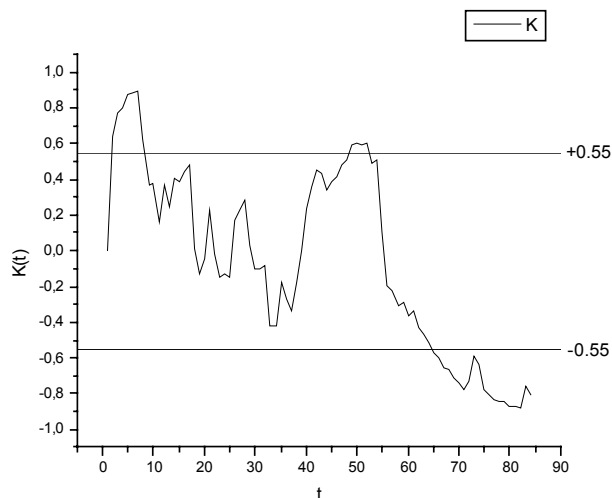


Рис.4

При высоком уровне случайных флуктуаций исходного ряда значения экспоненциальной средней могут весьма медленно реагировать на изменение его среднего уровня в силу сильной зависимости от предистории процесса. Установлено, что для уменьшения времени запаздывания реакции может быть успешно использован разработанный Р.Брауном и Д.Триггом адаптивный фильтр экспоненциального сглаживания, параметры которого формируются на основе так называемого следящего сигнала [3]. Показано, что по выходу значения следящего сигнала K_t за соответствующие контрольные границы можно приблизительно определять время изменения тенденции показателя посещаемости (рис. 4).

В итоге разработана и опробована на конкретных примерах методика решения комплекса задач статистической оценки тенденций посещаемости web-сайтов, которая может использоваться в дополнение к известным коммерческим программным продуктам.

ЛИТЕРАТУРА:

1. Линник Ю.В. Метод наименьших квадратов и основы теории обработки наблюдений. Москва, Физматгиз, 1962. – 312 с.
2. Фёрстер Э., Рёнц Б. Методы корреляционного и регрессивного анализа. Москва, “Финансы и статистика”, 1983. – 301 с.
3. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования. Москва, “Статистика”, 1979. – 254 с.