

ОПРЕДЕЛЕНИЕ ЭФФЕКТИВНОЙ СТРУКТУРЫ ВХОДНЫХ ДАННЫХ ДЛЯ ОБУЧЕНИЯ НЕЙРОННОЙ СЕТИ РЕШЕНИЮ ЗАДАЧИ ПРОГНОЗИРОВАНИЯ СПРОСА

Грицай А.А.

Кафедра информационных технологий

Поступила в редакцию 11.06.2014, после переработки 16.06.2014

Основная цель работы состоит в исследовании различных методов определения эффективной входной структуры данных для нейронных сетей применительно к задаче прогнозирования спроса, а также разработке на их основе более точного алгоритма с более широким спектром применимости. Это позволит автоматизировать процесс создания обучающих примеров из временного ряда. Критерием точности выступает мера ошибки прогнозирования sMARE. Основным результатом работы является разработанный алгоритм ранжирования, представляющий собой композицию рассмотренных методов. В работе показана эффективность предлагаемого алгоритма, приведены результаты тестирования на реальных данных о спросе на продукцию оптово-розничных торговых предприятий.

Ключевые слова: нейронные сети, прогнозирование спроса, структура нейронной сети, оптимизация запасов.

Вестник ТвГУ. Серия: Прикладная математика. 2014. № 2. С. 95–106.

Введение

В статье рассматривается алгоритм автоматического выбора эффективной структуры входных данных нейронной сети. Под структурой входных данных понимается отображение исходного обучающего множества в набор обучающих примеров для настройки параметров авторегрессионной модели. Эффективность структуры входных данных оценивается по ошибке прогнозирования, полученной на авторегрессионной модели с такой структурой входных данных. В качестве критерия точности прогнозирования используется мера ошибки sMARE. Рассматриваемый алгоритм представляет из себя часть разрабатываемой системы управления запасами Forecast NOW! (fnow.ru) Система решает задачу оптимизации товарных запасов оптово-розничных торговых предприятий. Разработка такого алгоритма является актуальной, так как он позволяет снизить требуемый уровень квалификации пользователя при использовании нейронных сетей для решения задачи прогнозирования спроса на продукцию и определения оптимального запаса. Алгоритмы проверялись на базе данных реальных торговых предприятий, число товарных позиций в которой составляет порядка 20 тыс. единиц, размер истории

продаж составляет как минимум два года по каждой позиции. Рассматриваемые торговые предприятия осуществляют деятельность по направлению торговли автозапчастями, продуктами питания, хозяйственными товарам, электротоварами и стройматериалами. В исследовании участвуют позиции, имеющие частый спрос, т.е. средний интервал между двумя фактами продаж которых не превышает 1.25 периода. Это позволяет говорить об эффективности применения авторегрессионной модели [5]. В качестве регрессионной модели в рамках данного исследования используется нейронная сеть. На вход нейронной сети подаются значения прошлого спроса (восстановленного из истории продаж, остатков, проводимых промоакций и другой дополнительной информации). Алгоритм позволяет выбрать наиболее оптимальную глубину погружения в лаговое пространство и оценить, какие значения спроса в прошлом наилучшим образом отражают будущие тенденции и в большей степени содержат информацию о нем.

1. Постановка задачи

Перед нами стоит задача прогнозирования временных рядов. Для решения этой задачи используется нейронная сеть.

Пусть временной ряд задается следующим образом

$$X = \{x(i) : 0 \leq i \leq n\}.$$

Размерность рассматриваемого лагового пространства (пространства вложения) равна d , а лаг равен τ . Тогда, мы можем построить следующие лаговые векторы $X_i = [x(i), x(i + \tau), \dots, x(i + (d - 1)\tau)]$. Согласно теореме Такенса [2] для достаточно большой размерности лагового пространства d , лаговые векторы позволяют однозначно точно восстановить фазовое пространство с идентичными свойствами исходной системы. Таким образом, требуется определить минимальную размерность пространства вложения d . Задача прогнозирования таким образом сводится к задаче распознавания образов [4].

Дадим определение структуры входных данных.

Определение 1. Структура входных данных есть отображение исходного временного ряда в набор обучающих примеров (векторов размерности $d + 1$). Каждый обучающий пример состоит из входа (вектор $X_i = [x(i), x(i + \tau), \dots, x(i + (d - 1)\tau)]$) и ожидаемого выхода ($Y_i = x(i + 1)$).

Определение 2. Эффективная структура входных данных есть структура входных данных, использование которой позволяет достигнуть субоптимального решения относительно ошибки прогнозирования.

Задача, решаемая в ходе данного исследования, состоит в разработке алгоритма автоматического поиска эффективной структуры входных данных без необходимости полного перебора или обучения системы на каждой такой структуре. Для оценки эффективности методом обратного распространения ошибки обучалась нейронная сеть типа многослойный перцептрон. На вход сети подавались обучающие примеры в соответствии с определенной алгоритмом эффективной структурой входных данных. В качестве меры эффективности используется мера ошибки SMAPE, впервые рассмотренная в работе [6].

$$SMAPE = \frac{1}{n} \sum_{t=1}^n \frac{|F_t - A_t|}{A_t + F_t},$$

где F_t – прогнозируемое значение временного ряда в момент времени t , A_t – фактическое значение временного ряда в момент времени t , n – длина временного ряда.

2. Решение задачи

В статье были реализованы и исследованы следующие известные алгоритмы решения задачи:

1. Алгоритм совместной информации [7].
2. Метод на основе вычисления функции автокорреляции [8].
3. Оценка совместности обучающего набора [10].
4. Модифицированный алгоритм k-ближайших соседей для поиска ошибочных ближайших соседей при отображении [9].

По результатам исследования были разработаны следующие алгоритмы для решения задачи поиска эффективной структуры входных данных для нейронной сети: алгоритм ранжирования и алгоритм приращения энтропии (разработан на базе алгоритма совместной информации).

Алгоритмы, если это было необходимо, модифицировались для получения конкретного значения глубины погружения в лаговое пространство. Каждый метод выдает вектор возможности, с которой для данного временного ряда является оптимальным тот или иной лаг. Каждый алгоритм был протестирован на наборе синтетических и реальных данных. Предположений о статистической природе выборки не делалось. Реальные данные были представлены временными рядами продаж товаров с различными характерными особенностями:

1. ярко выраженная недельная сезонность,
2. наличие другой краткосрочной сезонности
3. отсутствие краткосрочной сезонности
4. наличие/отсутствие тренда,
5. товары с высокой/низкой частотой продажи
6. большие/низкие объемы среднесуточных продаж
7. неслучайное отсутствие спроса,
8. ограниченный спрос в условиях дефицита товара.

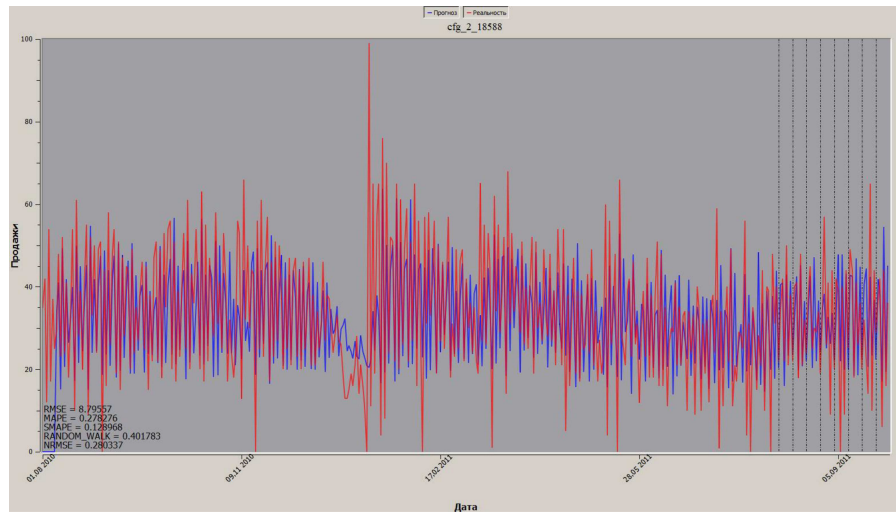


Рис. 1: Высокочастотные продажи, наличие недельной и годовой сезонности

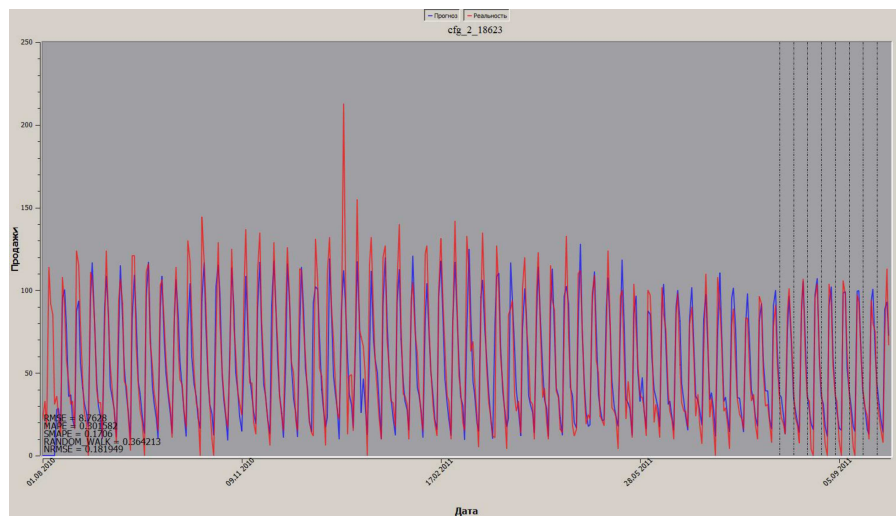


Рис. 2: График продаж товара «газета» с ярковыраженной недельной сезонностью

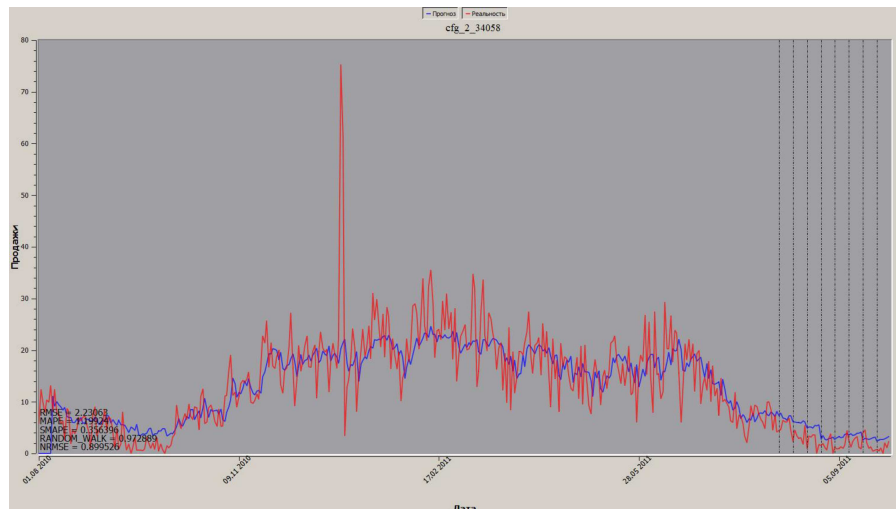


Рис. 3: График продаж товара с ярковыраженной годовой сезонностью

В качестве примеров, иллюстрирующих исходную выборку, рассмотрим некоторые временные ряды и построенные по ним прогнозы. На Рис. 1, 2, 3 красным представлены реальные продажи, синим – построенные прогнозы.

Синтетические данные включали в себя временные ряды из набора данных The Santa Fe Institute Time Series Prediction and Analysis Competition. Подробнее о данном наборе данных и соревновании по прогнозированию временных рядов можно прочитать в работе [4].

Рассмотрим известные методы для определения эффективной структуры входных данных.

2.1 Алгоритм на основе совместной информации

Совместная информация [3] с задержкой во времени – инструмент для определения адекватной глубины погружения в лаговое пространство. В отличие от автокорреляционной функции, совместная информация может также учитывать нелинейные зависимости. Вычисляется согласно формуле

$$S = - \sum_{ij} p_{ij}(t) \ln \frac{p_{ij}(t)}{p_i p_j},$$

где p_i – вероятность найти значение временного ряда в i -ом интервале, $p_{ij}(t)$ – совместная вероятность события, что наблюдаемое значение одновременно попадет на i -ый интервал и наблюдаемое значение с лагом t попадет в j -ый интервал. Определяется значение, при котором совместная информация достигает наперед заданного минимума. Лаг, при котором происходит это событие, считается оптимальным. Данный метод был реализован и протестирован на реальных данных. Метод показал результат в 77,2%. Пример графика зависимости совместной информации от лага представлен на Рис. 4.

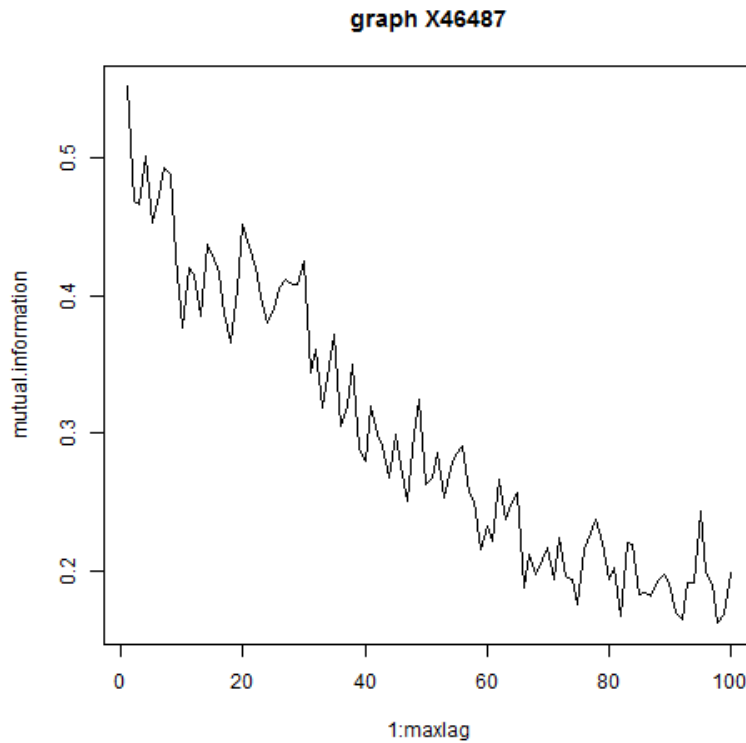


Рис. 4: График зависимости совместной информации от лага

2.2 Алгоритм на основе вычисления автокорреляционной функции

Автокорреляция показывает степень линейной статистической зависимости между исходным рядом и рядом, полученным из исходного, с определенным лагом. Значение автокорреляционной функции для лага τ определяется по формуле

$$R(\tau) = \frac{E[(X_t - \mu)(X_{t+\tau} - \mu)]}{\sigma^2}.$$

Все значения $R(\tau)$ меньше 0.1 признаются незначимыми. Согласно разработанному алгоритму, оптимальный лаг определяется как τ , при котором достигается $\max_{\tau} R(\tau)$, при этом рассматриваются τ не более чем наперед заданная константа. На Рис. 5 представлен пример зависимости значения функции автокорреляции от лага.

2.3 Алгоритм оценки совместности обучающего набора

Набор называется совместным [10], если различным входным частям обучающих примеров соответствуют различные ответные части. Таким образом, может

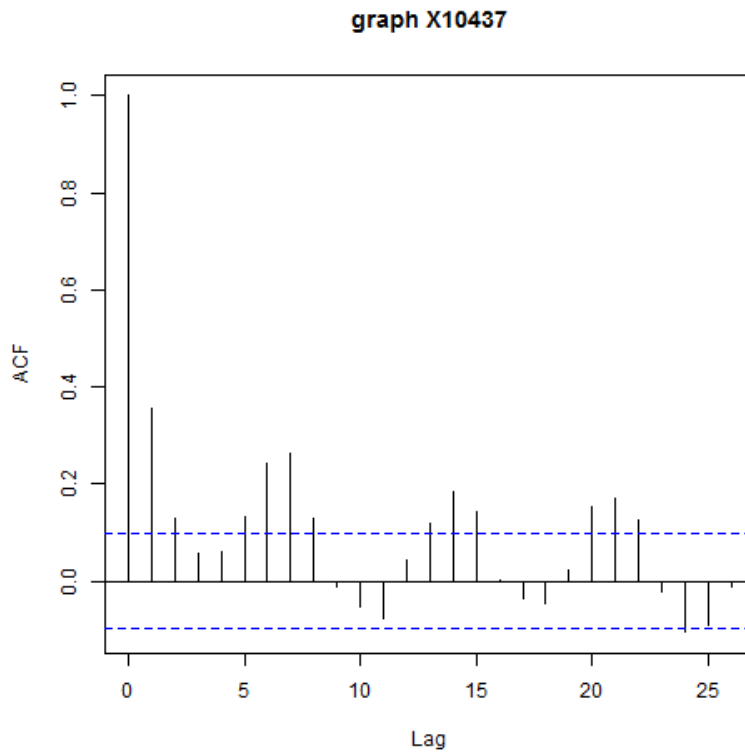


Рис. 5: График зависимости значения функции автокорреляции от лага

быть оценена «сложность» полученной обучающей выборки. Для вычисления степени несовместности набора применяется следующая формула

$$L = \max_{i \neq j} \frac{\|y^i - y^j\|}{\|x^i - x^j\|},$$

где $x^i \in R^n$, $y^i \in R^m$ – входные и выходные вектора соответственно

Было установлено, что оценка несовместности (противоречивости) набора, посчитанная согласно формуле ниже, дает более точные результаты и устойчива к выбросам:

$$L = \frac{1}{N} \sum_{i \neq j} \frac{\|y^i - y^j\|}{\|x^i - x^j\|}.$$

2.4 Модифицированный алгоритм k -ближайших соседей для поиска ошибочных ближайших соседей

Алгоритм ошибочных ближайших соседей (false nearest neighbor) метод для определения оптимальной размерности пространства при проекции исходного вре-

менного ряда. Алгоритм ошибочных ближайших соседей был впервые описан в работе [1].

Алгоритм ошибочных ближайших соседей позволяет определить минимальную размерность пространства вложения d .

Идея алгоритма следующая. Для каждой точки \vec{y}_i во временном ряде ищется ближайший сосед \vec{y}_j в k -размерном пространстве. Точка считается ближайшим соседом, если расстояние $\|\vec{y}_i - \vec{y}_j\|$ является минимумом среди всех других точек. Итерация происходит по обеим точкам и вычисляется

$$R_i = \frac{|x_{i+d\tau} - x_{j+d\tau}|}{\|\vec{y}_i - \vec{y}_j\|}.$$

Если R_i превышает заранее заданный эвристический порог R_t , то эта точка помечается как имеющая ошибочного ближайшего соседа. Размерность пространства достаточна для непротиворечивого проецирования, если доля точек для которых $R_i > R_t$ достаточно мала (меньше заданного порога).

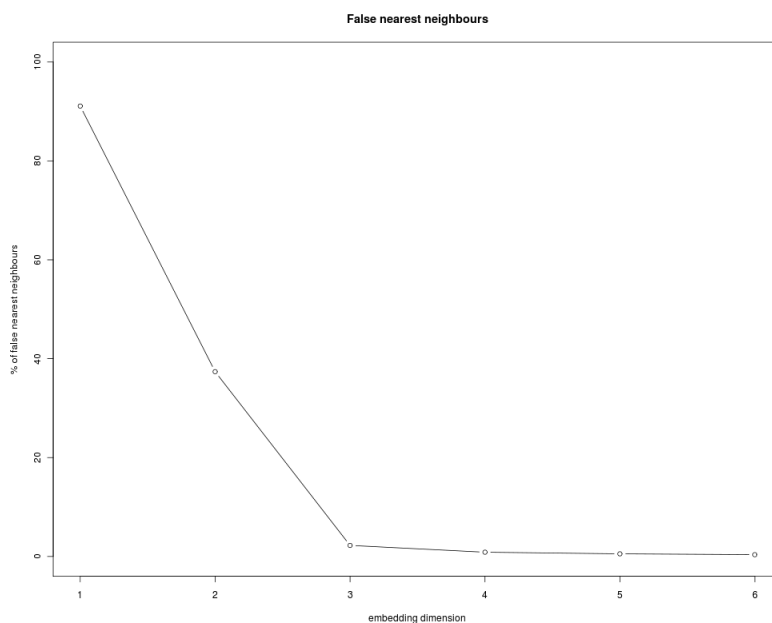


Рис. 6: График зависимости процента ошибочных ближайших соседей от лага

На основе существующих алгоритмов были разработаны алгоритм на основе вычисления приращения энтропии и метод ранжирования.

2.5 Алгоритм на основе вычисления приращения энтропии

Алгоритм приращения энтропии является модификацией алгоритма совместной информации. После вычисления совместной информации на различных лагах

происходит вычисление разностей первого порядка. Далее считается накопительная сумма по этим разностям, находится точка, в которой накопительная сумма перестает увеличиваться значительно (наперед заданный процент) или начинает значительно убывать. Также рассматривалась модификация, при которой остановка происходит по относительному приращению значения совместной информации относительно текущего максимума. Данное усовершенствование алгоритма позволило повысить точность системы распознавания на 1.7%. Метод показал результат в 78.9%. Согласно кластеризации метод хорошо работает на одних и тех же временных рядах, что и метод подсчета совместной информации. Таким образом, не является целесообразным включение алгоритма совместной информации в алгоритм ранжирования. На Рис. 7 представлен график зависимости приращения энтропии от лага для одного из временных рядов реальных продаж.

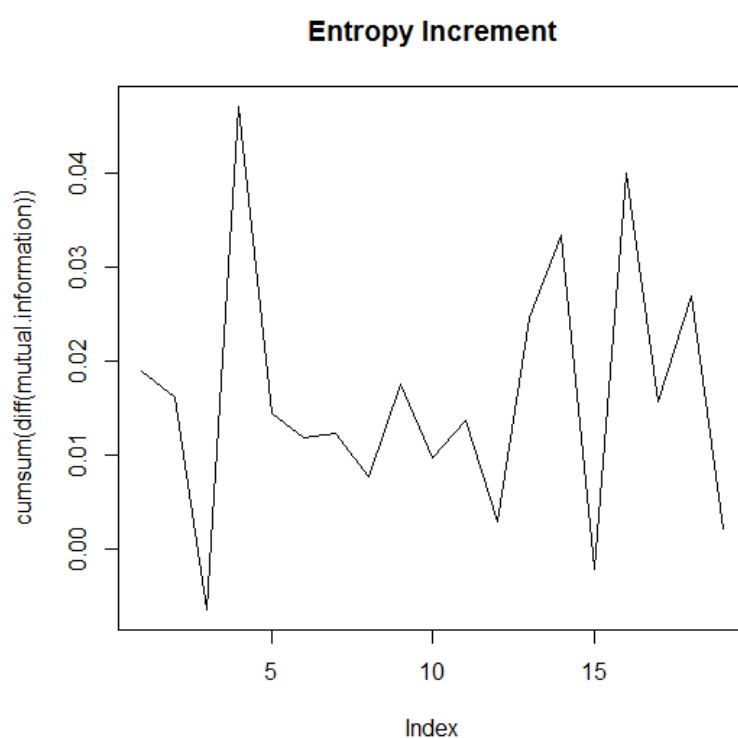


Рис. 7: График зависимости приращения энтропии от лага

2.6 Метод ранжирования

На базе реализованных алгоритмов разработан алгоритм ранжирования результатов от всех рассмотренных методов. Суть алгоритма заключается в следующем. Каждому методу назначается коэффициент доверия, а результирующий

Таблица 1: Точность системы прогнозирования при автоматическом выборе структуры входных данных различными методами

Метод	Средняя точность прогнозов по реальным данным
Совместной информации	77,2%
Автокорреляции	74%
Оценки совместности обучающего набора	73,3%
Модифицированный алгоритм k-ближайших соседей для поиска ошибочных ближайших соседей при отображении	76,5%
Метод приращения энтропии	78,9%
Метод ранжирования	81,3%

вектор представляется как взвешенная сумма в соответствии с полученными векторами и назначенными коэффициентами. Коэффициенты доверия выбираются таким образом, чтобы минимизировать ошибку прогнозирования по всей обучающей выборке.

Каждый метод обладает достоинствами в отношении определения одних зависимостей (например, линейных в случае автокорреляции), а в случае других видов зависимостей оказывается неэффективным. Кластеризация по ошибке показала, что существуют кластеры временных рядов, для которых одни методы выдают результаты лучше, чем другие. Значит, использование композиции алгоритмов представляется целесообразным. В Таблице 1 представлены результаты точности системы прогнозирования при использовании каждого метода в отдельности, а также разработанного метода ранжирования. По ней можно видеть, что разработанный в статье метод ранжирования позволяет увеличить точность прогнозирования (до 81,3%) за счет учета векторов возможностей различных алгоритмов и применения сильных сторон тех или иных алгоритмов в специфичных случаях согласно кластерному анализу.

Заключение

В ходе проведенных исследований и разработок был реализован алгоритм автоматического выбора эффективной структуры данных для решения задачи прогнозирования с использованием нейронных сетей. Предлагаемый алгоритм обладает большей точностью и позволяет снизить необходимый уровень квалификации пользователя для применения нейронных сетей в качестве инструмента прогнозирования спроса. Этот факт является немаловажным в контексте решаемой нами задачи автоматизации процесса управления запасами торговых предприятий.

Список литературы

- [1] Kennel M.B., Brown R., Abarbanel H.D.I. Determining embedding dimension for phase space reconstruction using the method of false nearest neighbours // Physical Review A - Atomic, Molecular, and Optical Physics. 1992. Vol. 45, № 6. Pp. 3403–3411.
- [2] Takens F. Detecting strange attractors in turbulence. Dynamical Systems and Turbulence. Springer: New York, 1980. Pp. 366–381.
- [3] Peng H., Long F., Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2005. Vol. 27, № 8. Pp. 1226, 1238.
- [4] Weigend A.S., Gershenfeld N.A. Time series prediction: Forecasting the future and understanding the past. Boston: Addison-Wesley, 1994.
- [5] Syntetos A.A., Boylan J.E. The accuracy of intermittent demand estimates // International Journal of Production Economics. 2005. Vol. 21. Pp. 303–314.
- [6] Armstrong J.S. Long-range Forecasting: From Crystal Ball to Computer. Wiley, 1978.
- [7] Ai-Hua Jiang, Xiu-Chang Huang, et al. Mutual information algorithms // Mechanical Systems and Signal Processing. 2010. Vol. 24. Pp. 2947–2960.
- [8] Broze L., Francq C., Zakoian J-M. Efficient use of higher-lag autocorrelations for estimating autoregressive processes // Journal of Time Series Analysis. 2002. Vol. 23, № 3. Pp. 287–312.
- [9] Rhodes C., Morari M. The false nearest neighbors algorithm: An overview // Computers and Chemical Engineering. 1997. Vol. 21. Pp. S1149–S1154.
- [10] Тарасенко Р.А., Крисиллов В.А. Предварительная оценка качества обучающей выборки для нейронных сетей в задачах прогнозирования временных рядов // Труды Одесского политехнического университета. 2001. № 1(13). С. 90–93.

Библиографическая ссылка

Грицай А.А. Определение эффективной структуры входных данных для обучения нейронной сети решению задачи прогнозирования спроса // Вестник ТвГУ. Серия: Прикладная математика. 2014. № 2. С. 95–106.

Сведения об авторах**1. Грицай Александр Александрович**

аспирант кафедры информационных технологий Тверского госуниверситета.

Россия, 170100, г. Тверь, ул. Желябова, д. 33, ТвГУ, ПМуК.

E-mail: allgrit@forecastnow.ru.

DETERMINATION OF THE OPTIMAL INPUT DATA STRUCTURE FOR TRAINING THE DEMAND FORECASTING NEURAL NETWORK

Gritsai Alexander Alexandrovich

Postgraduate student of Information Technology chair, Tver State University
Russia, 170100, Tver, 33 Zhelyabova str.
E-mail: allgrit@forecastnow.ru

Received 11.06.2014, revised 16.06.2014.

The main objective of the work is to study the various methods of determining the optimal structure of the input data for the neural networks applied to the problem of forecasting demand, as well as on the development of more accurate algorithm with a wider range of applicability. It will automate the process of creating training examples from the time series. The criterion for the accuracy of the prediction error measure is sMAPE. The main result is the developed ranking algorithm, which is a composition of the methods considered above. This paper shows the effectiveness of the proposed algorithm, the results of tests on real demand time series.

Keywords: neural networks, demand forecasting, neural network structure, inventory optimization.

Bibliographic citation

Gritsai A.A. Determination of the optimal input data structure for training the demand forecasting neural network. *Vestnik TverGU. Seriya: Prikladnaya matematika* [Herald of Tver State University. Series: Applied Mathematics], 2014, no. 2, pp. 95–106. (in Russian)