

І.В. Скрипіна, ст. викл.

Харківський національний автомобільно-дорожній університет

МЕТОДИ ПРОГНОЗУВАННЯ ПАСАЖИРОПОТОКУ НА МЕТРОПОЛІТЕНІ

(Представлено д.т.н., проф. Панішевим А.В.)

Наведено огляд математичних методів, за допомогою яких можливо спрогнозувати кількість перевезених пасажирів на метрополітені. Для прогнозування пропонується нейромережева модель.

Аналіз існуючих методів прогнозування

Ретельний і всебічний аналіз при виборі методів прогнозування спрощує процедуру прийняття рішення. Вибраний метод повинен включати в себе складний математичний процес і давати точний, своєчасний і зрозумілий прогноз.

За оцінками дослідників число різних методів, прийомів і методик прогнозування перевищило 160 [4]. Однак число базових методів нараховує декілька десятків. Специфіка початкової інформації та об'єкта прогнозування вимагає вибору адекватного методу прогнозування. Під методом прогнозування розуміють сукупність прийомів прогнозування, що дозволяють на основі ретроспективних даних, відомих зовнішніх та внутрішніх зв'язків підсистем, а також їх змін, вивести думки певної достовірності відносно майбутніх показників. Зі всієї безлічі моделей можна виділити 5 основних груп: евристичні, математичні методи тимчасової екстраполяції, математичні методи просторової екстраполяції, методи моделювання процесів розвитку, логічні та структурні методи штучного інтелекту.

Специфіка початкової інформації та об'єкта прогнозування вимагають адекватного методу. Отримана нами інформація є набором даних спостережень, зібраних через послідовні проміжки часу. Іншими словами, є тимчасовий ряд, для якого необхідно вибрати відповідний метод прогнозування. У даній статті розглянемо математичні методи.

Існують дві основні цілі аналізу тимчасових рядів – це визначення природи ряду і прогнозування. Обидві цілі вимагають, щоб модель ряду була ідентифікована і формально описана.

За своєю природою тимчасовий ряд може відповідати різним моделям поведінки даних в наборі. Існує чотири основних типи моделей даних: горизонтальна, тренд, сезонна та циклічна.

Основним фактором, що впливає на вибір методу прогнозування, є ідентифікація реальних моделей, присутніх в даних.

Для стаціонарних даних використовуються ковзаючі середні [1], просте експонентне згладжування [2] і методи авторегресійного ковзаючого середнього (методи Бокса–Дженкінса) [3]. Для рядів, що мають тренд, метод лінійного експонентного згладжування Хольта, криві, що зростають, експонентні моделі і методи авторегресійних інтегрованих ковзаючих середніх (методи Бокса–Дженкінса). Для аналізу рядів, що мають явно виражену сезонність, використовують класичне розкладання, метод Census X-12 [5], експонентне згладжування Вінтера, багатомірну регресію тимчасового ряду і методи Бокса–Дженкінса. Для аналізу циклічних рядів необхідно включити економічні індикатори та економетричні моделі.

Для вибору методу прогнозування на першому етапі дослідження будемо використовувати дані про вхідний платний пасажиропотік на найбільш завантаженій лінії "Холодна гора – Пролетарська". Вибір платного пасажиропотоку для випробування методів прогнозування пояснюється тим, що дані про пільговий пасажиропотік менш точні, ніж дані про платні пасажиропотоки. Вибір лінії обумовлений присутністю на ній приблизно однакових для всіх станцій закономірностей коливання пасажиропотоку по місяцях року і днях тижня, кожній станції і лінії загалом властиве зростання кількості пасажирів, яку можна апроксимувати лінійним або нелінійним рівнянням тренду.

З урахуванням специфіки пасажиропотоку на першій лінії для аналізу були відібрані шість варіантів побудови моделей тимчасових рядів.

Варіант 1. Прогнозування методом експонентного згладжування в припущенні, що тимчасовий ряд містить лінійний тренд і сезонність (коливання по днях тижня) адитивного типу.

Варіант 2. Прогнозування методом експонентного згладжування в припущенні, що тимчасовий ряд містить лінійний тренд і сезонність (коливання по днях тижня) мультипликативного типу.

Варіант 3. Прогнозування методом експонентного згладжування в припущенні, що тимчасовий ряд містить нелінійний тренд, виражений рівнянням зростання з ефектом насичення (сповільнення темпів зростання), і сезонність (коливання по днях тижня) адитивного типу.

Варіант 4. Прогнозування методом експонентного згладжування в припущенні, що тимчасовий ряд містить нелінійний тренд, виражений рівнянням зростання з ефектом насичення (сповільнення темпів зростання), і сезонність (коливання по днях тижня) мультипликативного типу.

Процедура побудови моделі для кожного з перших чотирьох варіантів була реалізована за допомогою пакета STATISTICA 6.0.

Варіант 5. Прогнозування з використанням моделі нейронної мережі.

Варіант 5 реалізовувався з використанням генетичного алгоритму пакета STATISTICA Neural Networks [3]. Оцінка якості моделі проводилася за величиною середньої відносної помилки прогнозу даних спостережень. Порівняльна таблиця результатів наведена нижче у табл. 1.

Таблиця 1

Помилки прогнозів для різних варіантів моделей

Варіанти моделей	Характеристика моделі	Середня квадратична, люд.	Середня відносна, %
Варіант 1	Exp. smoothing: Additive season (7) S0=147E3 T0=110,8 Lin.trend, add.season; Alpha=, 100 Delta=, 100 Gamma=, 100	24776	8,42
Варіант 2	Exp. smoothing: Multipl. season (7) S0=147E3 T0=110,8 (line_1) Lin.trend, mult.season; Alpha=, 254 Delta=0,00 Gamma=0,00	23244	7,47
Варіант 3	Exp. smoothing: Additive season (7) S0=144E3 T0=1108, (line_1) Damped trend, add.season; Alpha=, 254 Delta=0,00 Phi=, 100	23296	7,34
Варіант 4	Exp. smoothing: Multipl. season (7) S0=144E3 T0=1108, (line_1) Damped trend, mult.season; Alpha=, 254 Delta=0,00 Phi=, 100	23589	7,39
Варіант 5	MLP (модель багатошарового персептрона)	21591	7,27

З таблиці видно, що найкращою є модель, яка використовує нейронну мережу багатошарового персептрона. Середня величина відносної помилки для цієї моделі рівна 7,27 %.

Пошук кращої моделі нейронної мережі був виконаний з використанням IPS STATISTICA Neural Networks

Потрібно зазначити, що хороші результати роботи нейронна модель дає в середньому. Якщо звернутися до отриманих даних, то можна побачити, що вона недостатньо добре описує поведінку пасажиропотоку у вихідні і святкові дні, «спотикається» в оцінці пасажиропотоку днів, що включаються в так званий святковий цикл. Відносні помилки прогнозу в ці дні досягають до 20 %, 1 січня, 1 травня та в деякі інші святкові дні помилка досягає до 80 %. Тоді як помилка прогнозу в тижні, що не містить святкових днів і перенесень днів відпочинку, не перевищує 5 %.

У зв'язку з останнім висновком нами була розроблена модель нейронної мережі по укороченому тимчасовому ряду, з якого були виключені тижні, що містять святкові, цикли. Внаслідок цього середня відносна помилка прогнозу даних поменшала до 4,8 %, а розкид значень лежав в інтервалі від 0 до 15 %. Цей результат дозволяє зробити висновок про доцільність розробки двох методик прогнозування пасажиропотоків: для тижнів звичайних і тижнів, що містять святкові дні.

Розглянемо наступний варіант 6. Модель нейронної мережі із збільшеним числом шарів для тимчасового ряду, в якому виправлене значення пасажиропотоку в день проведення акції "безкоштовний проїзд в метро" 26.07.2001 року.

При розгляді результатів прогнозу по днях, отриманих за варіантом 5, виявилось, що 26.07.2001 року у звітності метрополітену по першій лінії був показаний вхідний платний пасажиропотік 21170 чоловік, тоді як фактичне значення звичайно на порядок більше і повинне було знаходитися в межах 200000...250000 чол. Причина розходжень – проведення акції "безкоштовний проїзд в метро". Ми замінили в цей день звітне значення на розрахункове.

Крім того, ми змінили стратегію пошуку кращої нейронної моделі: збільшили число типів моделей, а в деяких з них – число шарів нейронної мережі, переглянули співвідношення між кількістю даних в повчальній, тестуючій та контрольній вибірках.

Внаслідок роботи генетичного алгоритму кращою виявилася нейромережева модель з наступною структурою (рис. 1).

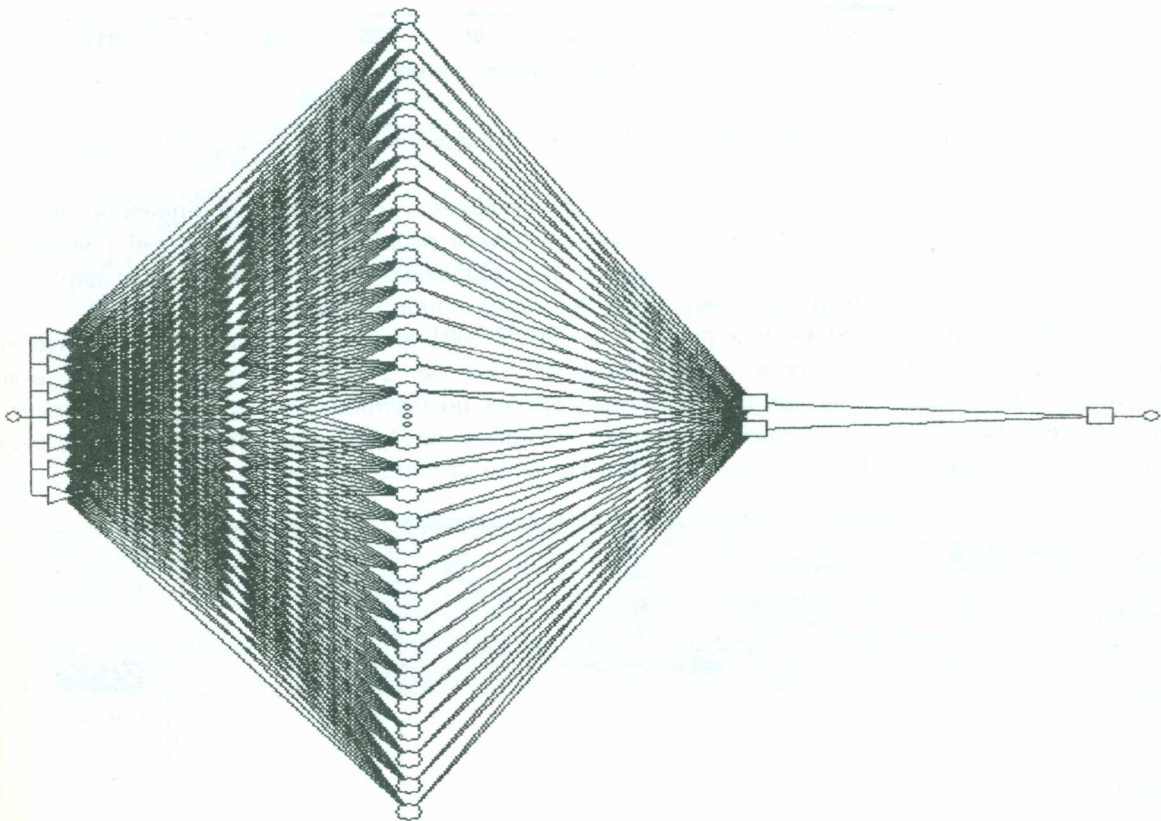


Рис. 1. Модель чотиришарової нейронної мережі, що найкращим чином описує тимчасовий ряд щоденного вхідного платного пасажиропотоку на лінії "Холодна гора – Пролетарська"

Середня відносна помилка прогнозу значень пасажиропотоку за цією моделлю рівна 3,6 %.

На рис. 2 наведена гістограма відносних відхилень передбачених даних від звітних.

Як видно, з 1064 спостережень в діапазон від 15 до +15 % попадає 1000 спостережень, і 64 спостереження виходять за його кордони. Аналіз розходжень показує, що найбільші розходження між звітними і передбаченими за моделлю значеннями припадають на вихідні та святкові дні.

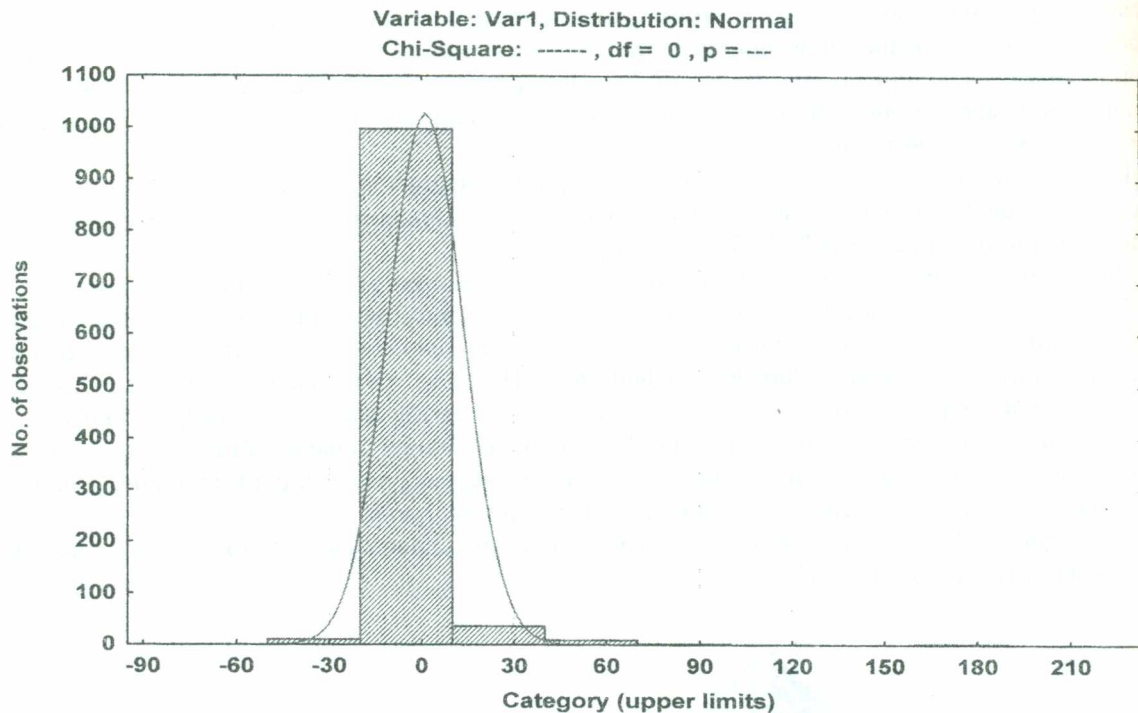


Рис. 2. Гістограма відхилень передбачених значень пасажиропотоку від звітних

Процес прогнозу може бути представлений графічно. Подальше поліпшення моделі можливе різними шляхами, наприклад, можна спробувати розділити ряд на типові і нетипові робочі тижні, розкласти рівні ряду на компоненти і моделювати кожну компоненту нарізно з подальшим узагальненням отриманих варіантів. Можливі й інші підходи.

Розглянемо варіант 7. Якщо у варіанті 6 по 26.07.2001 значення пасажиропотоку, що не враховується, 21170 було просто збільшено в 10 раз до 211700, то у варіанті 7 воно розраховане як середнє із значень цієї величини, взятих по п'ятниці в попередній і подальший тижні. У результаті середня відносна помилка поменшала до 2,7. Процес прогнозу може бути представлений графічно. На рис. 3 представлені графічні результати прогнозування.

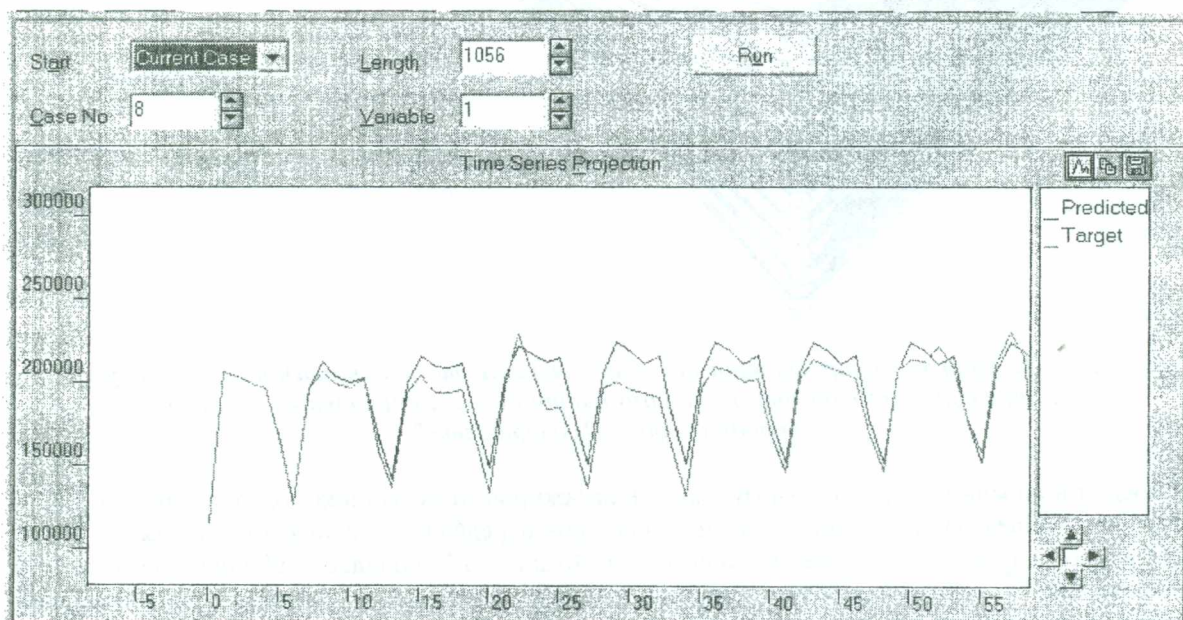


Рис. 3. Графічне представлення результатів прогнозування

Висновки

Аналіз розглянутих методів показав, що для прогнозування пасажиропотоку метрополітену на лінії, що розглядається, може бути запропонована нейромережева модель чотиришарового перцептрона.

Для більш ефективного прогнозу, крім закономірностей тимчасових рядів, необхідно враховувати також різні фактори, пов'язані з розвитком наземного транспорту, перспективами самого метрополітену, надалі перетворенням транспортної галузі загалом, що вказує про доцільність побудови багатфакторної моделі.

ЛІТЕРАТУРА:

1. *Андерсон Т.* Статистический анализ временных рядов. – М.: Мир, 1976. – 426 с.
2. *Бокс Дж., Дженкинс Г.* Анализ временных рядов. – М.: Мир, 1974. – Вып. 1. – 406 с.; Вып. 2. – 196 с.
3. *Боровиков В.В.* STATISTICA: искусство анализа данных на компьютере. Для профессионалов. – СПб: Питер, 2001. – 656 с.
4. *Perry S.* Applied Business Forecasting. - Management Accounting 72(3), 1994. – 40 с.
5. *Эндрю Ф.* Сигел. Практическая бизнес-статистика: Пер. с англ. – Издательский дом “Вильямс”, 2002. – 1056 с.

СКРИПНА Ірина Валентинівна – старший викладач кафедри інформатики Харківського національного автомобільно-дорожнього університету, пошукувач Житомирського державного технологічного університету.

Наукові інтереси:

- математичне моделювання;
- теорія розкладів та її застосування.

Подано 11.07.2004