

ШАНДЕР О.Е., аспірант (Українська державна академія залізничного транспорту)

Формування моделі прогнозування обсягів вагонопотоків на станціях залізничного полігону

Робота присвячена питанню розробки прогнозної моделі, яка б одночасно видавала достатньо точну інформацію про реальні обсяги вагонопотоків на станціях і час пред'явлення вагонів на відповідні станції. Сформовано архітектуру штучної нейронної мережі Елмана, яка містить два входи та два виходи і є основою створеної прогнозної моделі.

Ключові слова: парк вантажних вагонів, залізничний полігон, вагонопотоки, модель прогнозування, нейронні мережі.

Вступ та актуальність теми

Процеси в економіці України супроводжуються структурними змінами в усіх сферах господарювання, в тому числі - на залізничному транспорті. Тому в умовах ринкової економіки виникла ситуація, при якій виробники фактично не можуть чітко планувати обсяги виготовлення своєї продукції. Відповідно до цього плани перевезення вантажів не відповідають реальній потребі у вагонах. Це викликає певну невизначеність при перерозподіленні вагонного парку та організації вагонопотоків і призводить, або до нестачі, або до надлишку вагонів в певні моменти часу на певній станції та дирекції в цілому. А якщо врахувати збільшення функціонування операторських компаній (власників вагонів) на залізничному транспорті, що виступають як організатори та виконавці перевезень вантажів на комерційній основі, то актуальним є вирішення питань ефективної організації вагонопотоків на станціях полігону не залежно від власності вагонів, за умови виконання плану перевезень вантажів [3]. Виходячи з цього, за умови розподілу за маршрутами як навантажених так і порожніх вагонів [1, 2] в залежності від навантаження-вивантаження відповідних станцій важливим є прогнозування обсягів вагонопотоків на станціях залізничного полігону.

Постановка задачі

Таким чином, для ефективного використання створених моделей [1, 2] управління вагонним парком як залізниці так і операторських компаній, для автоматизованого планування роботи на залізничній мережі потрібно створити прогнозну модель для приведення точності вхідної інформації до належного рівня, а саме - кількість вагонів і час пред'явлення вагонів на відповідну станцію.

Одним із сучасних напрямків у створенні моделей автоматизованого прогнозування є використання математичного апарату штучних нейронних мереж. Саме штучні нейронні мережі є основним математичним апаратом, який використовується у рамках сучасного напрямку який має назву Datamining. Datamining - це глибинний автоматизований інтелектуальний аналіз даних. Прогнозні моделі, які використовують технології Datamining, дають можливість виявлення нетривіальних, непомітних навіть для ока експерта закономірностей при аналізі даних, дозволяють будувати високоякісні прогнози [5].

Вирішення задачі

Штучні нейронні мережі - математичні моделі, які побудовані за принципами біологічних нейронних мереж - мереж нервових клітин живих організмів. Штучні нейронні мережі представляють собою систему сполучених між собою простих процесорів - штучних нейронів. Ці штучні нейрони періодично отримують, обробляють і передають сигнали. Штучний нейрон можна розглядати як нелінійну функцію від одного аргументу - лінійної комбінації всіх вхідних сигналів. Дану функцію називають функцією активації або передатною функцією. На рис. 1 наведена типова схема штучного нейрона, який складається з наступних елементів: 1 - входи, на які поступають сигнали від інших нейронів, 2 - суматор вхідних сигналів, 3 - обчислювач передатної функції, 4 - виходи, з яких вихідний сигнал поступає на інші нейрони, $w_1, w_2 \dots w_i$ - ваги вхідних сигналів [6, 11].

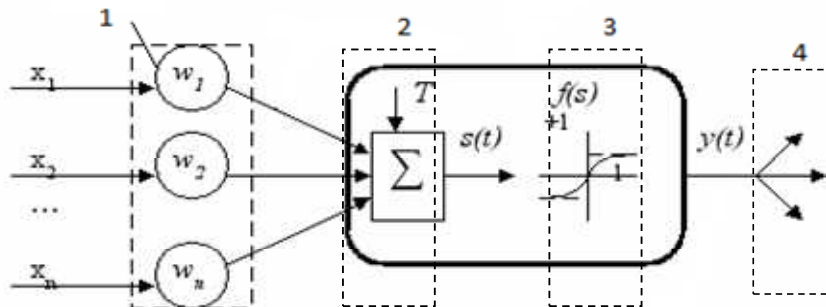


Рис. 1. Схематичне зображення штучного нейрона:

1 – входи, на які поступають сигнали від інших нейронів, 2 – суматор вхідних сигналів,
3 – обчислювач передатної функції,

4 – виходи, з яких вихідний сигнал поступає на інші нейрони, $w_1, w_2 \dots w_i$ – ваги вхідних сигналів [6,11].

Поняття штучних нейронних мереж тісно пов'язане з обробкою інформації за допомогою процесів навчання і адаптивного відгуку на вхідну інформацію. Нейромережі саме і використовуються у сфері здобуття, вивчення і узагальнення знань із масивів даних і образів. Основні класи задач, що вирішуються за допомогою штучних нейромереж – задачі класифікації і прогнозу. Є також багато видів нейромереж, які відрізняються методиками навчання, архітектурою: типами нейронів, кількістю шарів нейронів, видами зв'язків та інш. Тому для якісного вирішення конкретної задачі прогнозування потрібно обрати оптимальний тип і кількісні характеристики мережі.

Складність прогнозування часу подачі вагонів на станцію при замовленні клієнта і кількості відповідних вагонів полягає у тому, що ці два параметра пов'язані між собою і модель повинна оперувати ними одночасно враховуючі цей зв'язок при аналізі часових рядів даних.

Для побудови такої прогнозної моделі потрібно правильно обрати архітектуру штучної нейронної мережі.

Для вирішення поставленого завдання запропоновано використати динамічну нейронну мережу, яка має в своєму складі або елементи затримки або зворотні зв'язки і тоді стандартним вхідним сигналом мережі є послідовність векторів, які з'являються в певній часовій послідовності. Мережі які мають зворотні зв'язки також називають рекурентними [7].

Будову рекурентної нейронної мережі схематично зображено на рис. 2. Її входам та виходам відповідають шари рецепторних та ефекторних нейронів або просто лінії передачі даних. Між цими шарами розташовано один або кілька шарів прихованих нейронів. Входи нейронів кожного шару мають прямі зв'язки з

виходами нейронів попереднього шару та можуть мати зворотні зв'язки з виходами нейронів свого та наступних шарів. Зворотні зв'язки, як правило, містять елементи затримки, що надає нейромережі властивостей оперативної пам'яті. Нейрони різних шарів можуть бути однаковими або відрізнятися за типом активаційних функцій та характером нейропарадигми. Завдяки наявності затриманих зворотних зв'язків рекурентні мережі є динамічними системами, поведінка яких має зовнішню складову, що відповідає спостережуваним значенням входу та виходу, та приховану, яка характеризує внутрішній стан нейромережі. Ці складові поведінки представляють двома рівняннями, перше з яких є рівнянням вхід-вихід, а друге – рівнянням стану:

$$Y(t) = F[Z(t), X(t)], \quad (1)$$

$$Z(t) = \Phi[Z(t_0), X_{t_0}^t],$$

де $Y(t)$, $Z(t)$, $X(t)$ – значення відповідно виходу, стану та входу динамічної системи в момент t ;

$X_{t_0}^t$ – реалізація стимулу на вході системи в інтервалі часу $(t_0, t]$.

Однак і в динамічних мережах задавати і отримувати часові послідовності можна у різний спосіб [7]. У 1986 році американський вчений Майкл Джордан запропонував архітектуру динамічної нейронної мережі, яка використовує рекурентні зв'язки для створення динамічної пам'яті в мережі. Ця мережа мала прихований шар нейронів і рекурентні зв'язки виходу нейронів прихованого шару з його входами, що дало змогу нейронам прихованого шару “пам'ятати” свій стан на попередньому кроці. Цей метод був

розвинений в роботах іншого американського вченого – професора каліфорнійського університету Джефрі Елмана.

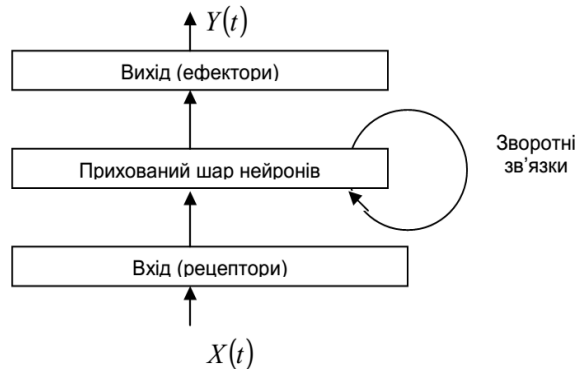


Рис. 2. Загальна структура рекурентної нейронної мережі

Питання про те, як представити час в рамках коннекціоністської моделі, тобто моделі, яка представляє собою мережу зв'язаних між собою простих елементів, що і представляє собою штучна нейронна мережа, дуже важливе. Один спосіб – представити час опосередковано, відображаючи його вплив на процеси, інший спосіб – явна інтерпретація часу (за допомогою моделі простору) [8]. Мережа Джордана була здатна вирішувати цікаві задачі, які поєднували в собі механізм вирішення задачі і механізм використання пам'яті.

Перший очевидний спосіб представлення зразків даних – проведення асоціації послідовного порядку даних з розмірністю вектора даних. Перший елемент відповідає першому моменту часу, другий – другому, і так далі. Такий підхід використовувався в багатьох попередніх моделях в яких використовувалася паралельна обробка даних. Але такий підхід має свої проблеми. Перш за все, така система повинна мати якийсь буфер між вхідним потоком і обчислювальним механізмом. По-друге, це також і логічна проблема: система повинна визначати коли потрібно використовувати буфер. Також проблемою є і розмір буфера. Він повинен забезпечувати розміщення зразка даних повністю, з чого можна зробити висновок, що всі вхідні вектори повинні мати однакову розмірність. Ще один дуже серйозний недолік такого підходу: при геометричній інтерпретації даних дуже важко визначити різницю між абсолютною і відносною позицією часової відстані між даними. Наприклад при такому підході 2 ідентичних зразка даних, які відрізняються лише невеликим здвигом у часі, фактично виглядають як абсолютно різні і віддалені у просторі. Систему звичайно можна навчити розпізнавати такі образи як подібні, але це буде результатом зовнішнього втручання (навчання), а не

подібності структури самих зразків. Ця подібність не є узагальнюючою: вона не розповсюджується на нові подібні зразки даних, які будуть пред'явлені системі.

Системи, які базуються на штучних нейронних мережах і які не мають вище перелічених недоліків, і були запропоновані.

Вперше Джорданом і було запропоновано підхід, який оперував часом не як розмірністю вхідного вектора, а зовсім інакше: як властивістю динамічної системи, тобто системи з пам'яттю. Елман модифікував систему Джордана. Її вигляд зображено на рис. 3.

Принциповою відмінністю мережі Елмана від мережі Джордана є те, що контекстний шар сполучений зворотнім зв'язком не з виходом мережі а з виходом прихованого шару. Це дає змогу враховувати попередню історію процесу, що спостерігається, а також накопичувати інформацію і запам'ятовувати послідовності.

Контекстний шар – це додатковий набір елементів (нейронів), яким система керується зі сторони входу. На вхід такої мережі може подаватися як послідовність одиничних даних, так і векторів в залежності від типу вирішуваної задачі. Робота такої мережі виглядає просто: на вхід подається послідовність даних, кожен елемент даних слідує за попереднім через стандартний інтервал часу (нібито керується таймером). В момент часу t вхідний блок отримує перший елемент послідовності. Контекстний блок має ініціуюче значення, наприклад 0,5 (як середина інтервалу $[0\ 1]$). Далі вхідний блок і контекстний блок активують прихований і далі прихований блок транслює сигнал вперед, активуючи вихідний блок. Прихований шар також передає сигнал до контекстного блоку за допомогою зворотного зв'язку. В залежності від задачі, що вирішується, фаза навчання може бути присутня або відсутня в робочому циклі мережі. Якщо така фаза присутня – вихід мережі порівнюється з еталонним зразком даних і за допомогою механізму зворотного розповсюдження похибки послідовно змінюються сили зв'язків (ваги). Сила зворотного зв'язку залишається незмінною. На наступному кроці, що відповідає моменту часу $t+1$, процедура повторюється. На цьому кроці контекстний блок містить значення, які дорівнюють значенням, що містилися в прихованому блоці на попередньому кроці t .

На рис. 3 зображена архітектура нейронної мережі Елмана, яка має один вхід і один вихід. Але для вирішення задачі прогнозування параметрів замовлень на пред'явлення вагонів на станції на мережі прогнозна модель повинна здійснювати прогнозування двох параметрів: часу пред'явлення (подачі) вагонів клієнту на відповідну станцію і кількість вагонів. Крім того, потрібно врахувати інформацію про кількість замовлених вагонів, яку клієнт надав залізниці.

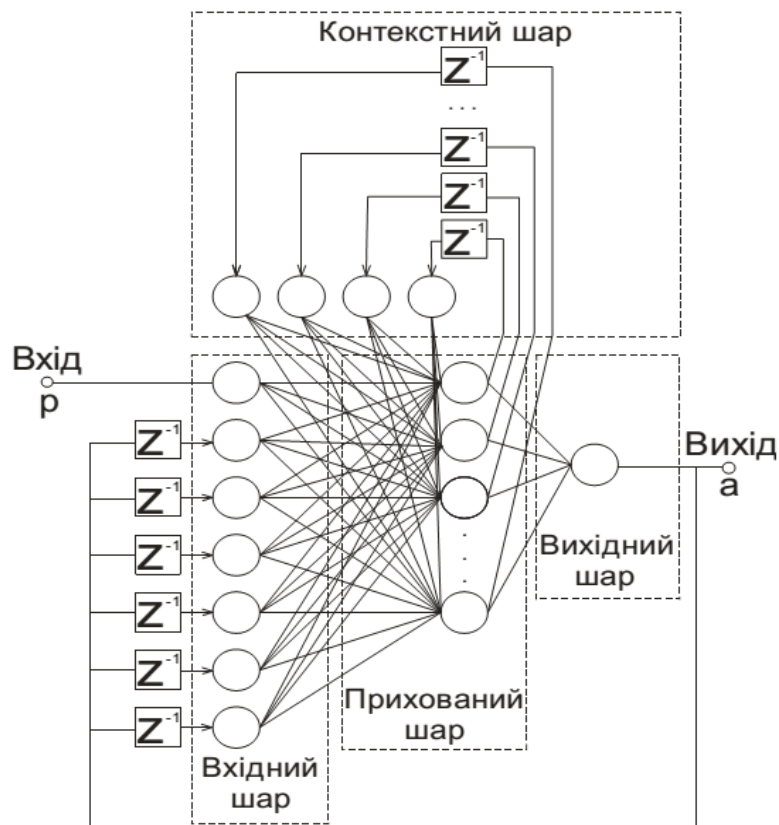


Рис. 3. Архітектура штучної нейронної мережі Елмана

Таким чином, запропонована прогнозна модель на базі нейронної мережі Елмана повинна містити два входи і два виходи. Перший вихід мережі призначений для виводу інформації про прогнозний час пред'явлення вагонів на станцію. Для урахування інформації про замовлену кількість вагонів доцільно прогнозувати не кількість вагонів, які клієнт буде в змозі завантажити, а різницю між замовленою кількістю вагонів і тією, яка за прогнозом буде використана фактично. Тоді другий вихід моделі представляє прогнозну різницю між замовленням і реальною потребою клієнта у вагонах на відповідній станції. На перший і другий входи моделі при навчанні і використанні моделі подаються історичні дані у вигляді часових рядів часів початку пред'явлення вагонів на станцію навантаження, які подані залізницею, і різниці між замовленою і фактично використаною кількістю вагонів відповідно. На рис. 4 наведено принциповий вигляд архітектури штучної рекурентної нейронної мережі Елмана, яка є основою створеної прогновної моделі, де X_1 , X_2 — нейрони вхідного шару, Y_1 , Y_2 — нейрони прихованого шару, u_1 , u_2 — нейрони контекстного

шару, Z_1 , Z_2 — нейрони вихідного шару, $W_{x_1y_1} \dots W_{y_1z_2}$ — ваги зв'язків між нейронами.

Кількість нейронів на кожному шарі показана мінімальною, оптимальна кількість нейронів, яка забезпечує задану точність прогнозу, визначається шляхом моделювання. Нейрони, які входять до рекурентного блоку (прихований і контекстний шари), мають функцію активації сигмоїдального типу, нейрони вхідного і вихідного шарів мають функцію активації лінійного типу.

Таким чином, для ефективного використання розробленої моделі розподілу парку вантажних вагонів для автоматизованого планування роботи на залізничній мережі, створили прогнозну модель для приведення точності вхідної інформації до належного рівня, а саме кількість вагонів і час пред'явлення вагонів на станцію залізницею під завантаження. Якщо розглядати модель управління парком вантажних вагонів операторських компаній то в якості прогнозних показників, тобто вхідною інформацією є також кількість вагонів та час пред'явлення вагонів операторською компанією на станцію.

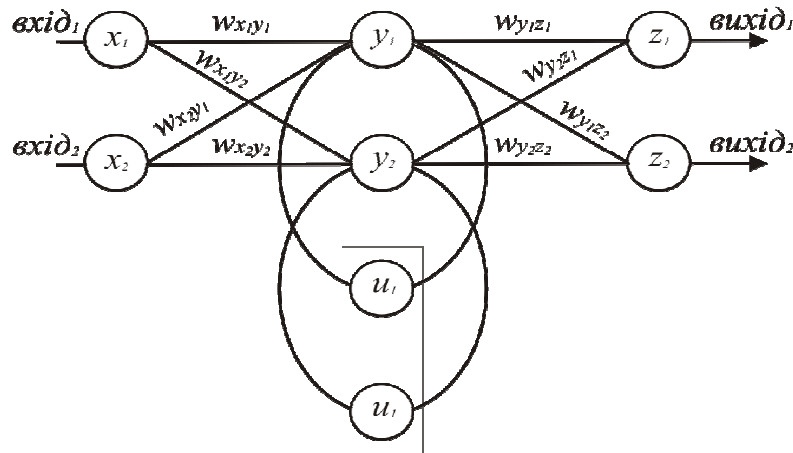


Рис. 4. Архітектура штучної нейронної мережі моделі прогнозування часу пред'явлення вагонів і кількості потрібних вагонів

Важливим та невід'ємним етапом побудови прогнозних моделей є верифікація.

Найбільш повною вважається інверсна верифікація, яка включає перевірку адекватності і точності прогнозної моделі за допомогою статистичних методів, методів прогнозного аналізу і прогнозної ретроспекції.

Адекватність моделі — узгодженість параметрів моделі і параметрів об'єкту або процесу, що моделюється. В нашому випадку параметри моделі — прогнозний час готовності пред'явлення вагонів на станцію і прогнозна кількість відповідних вагонів), а параметри процесу — фактичні значення цих параметрів, які мали місце після побудови прогнозу.

Для перевірки моделі на адекватність потрібно скористатися однією з спеціальних статистичних процедур. Однією з таких процедур є F-тест. Його суть полягає в тому, що висувається нульова гіпотеза про рівність дисперсій двох або більшого числа груп даних, наприклад прогнозні і фактичні дані, для яких обчислюється статистичний показник F-статистика, або критерій Фішера, і якщо цей показник не перевищує певного критичного значення, це означає, що він задовольняє розподіленню Фішера-Снедекора і нульова гіпотеза приймається, тобто ці групи даних можуть належати до однієї генеральної сукупності [4].

Якщо цих груп лише дві, тоді критерій Фішера обчислюється за наступною формулою:

$$F = \frac{\sigma_X^2}{\sigma_Y^2}, \quad (2)$$

де σ_X^2 , σ_Y^2 — дисперсії вибірок даних, при чому вибіркою X вважається та, значення дисперсії у якої більше.

Умовою прийняття нульової гіпотези є не перевищення критерієм Фішера критичного значення, що представляє собою значення зворотного двохпараметричного розподілу Фішера-Снедекора з параметрами (df_X, df_Y), які представляють собою статистичні ступені свободи наборів даних, тобто кількості незалежних випробувань, які можна обчислити як кількість елементів вибірки мінус один [4]. Це критичне значення відповідає значенню імовірності 0,05 (стандартний статистичний рівень значущості) зворотного розподілу Фішера-Снедекора і не може бути перевищене, якщо нульова гіпотеза не відкидається. Його можна обчислити за допомогою програмних пакетів, наприклад Microsoft Excel, або знайти у таблиці критичних значень F-критерію Фішера.

Для оцінки точності прогнозних моделей існує значна кількість критеріїв і показників. Більшість з цих показників представляють собою різні види абсолютних і відносних похибок прогнозу, які обчислюють на основі фактичних і прогнозних даних.

Одним із таких показників, який характеризує відносну точність прогнозу часових рядів, є середня абсолютна відсоткова похибка (Mean Absolute Percentage Error, MAPE) [10]

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|A_t - F_t|}{A_t} \cdot 100 \%, \quad (3)$$

де n - це кількість значень часового ряду,

A_t - поточне фактичне значення часового ряду,

F_t - поточне прогнозне значення часового ряду.

Ще одним показником, який вважається потужним вимірювачем точності прогнозних моделей, є середньоквадратична похибка (Root-Mean-Square Error, RMSE), який, на відміну від багатьох інших показників, є чутливим навіть до одиничних похибок прогнозу. Цей показник має ті ж одиниці виміру, як і прогнозні дані. Тобто він оцінює абсолютну похибку. Середньоквадратичну похибку можна обчислити за формулою [9]

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (A_t - F_t)^2}{n}} \quad (4)$$

Перевірка адекватності і точності моделі була проведена з використанням фактичних даних підприємств-клієнтів залізниці і прогнозних значень цих параметрів, отриманих за допомогою моделі на той же період планування. Для двох пар вибірок кількості вагонів і для двох пар вибірок часів початку подачі вагонів значення критерію Фішера не потрапило до критичної області, тому створену прогнозну модель можна вважати адекватною. Згідно до проведених розрахунків середньоквадратична похибка прогнозу для часів початку подачі складає 30-50 хвилин і є цілком задовільною з точки зору можливості використання такого прогнозу у якості вихідних даних при виконанні автоматизованого розрахунку плану. Середні значення кількості вагонів за фактичними даними і за даними прогнозу майже не відрізняються, це свідчить про відсутність зміщення прогнозу.

Висновки

На основі аналізу моделі короткотермінового прогнозування було зроблено висновок, що найбільш адаптованим методом, який дозволяє одночасно отримувати два вхідних параметра для відповідних моделей є динамічні нейронні мережі Елмана. Спираючись на це, в роботі сформовано архітектуру штучної нейронної мережі з двома входами і двома виходами. За даними верифікації сформована модель відноситься до класу високоточних, про що свідчить величина середньої абсолютної відсоткової похибки, яка не перевищує 10%.

Література

1. Шандер, О.Е. Формування процедури розподілу порожнього парку вантажних вагонів на залізничній мережі [Текст] / О.Е. Шандер // Інформаційно-керуючі системи на залізничному транспорті. - 2014. - № 5 С. 40-43.
2. Бутько, Т.В. Формалізація процесу управління парком вантажних вагонів операторських компаній [Текст] / Т.В. Бутько, О.Е. Шандер // Східно-Європейський журнал передових технологій. - 2014. - № 2/3(68). - С. 55-58.
3. Бутько, Т.В. Основні напрямки адаптації перевізного процесу на залізничному транспорті в умовах функціонування конкурентних транспортних компаній [Текст] / Т.В. Бутько, О.Е. Шандер // Збірник наукових праць УкрДАЗТ. - 2012.- вип. 131.- С. 26-30.
4. Lomax, Richard G. (2007) Statistical Concepts: A Second Course, p. 10, ISBN 0-8058-5850-4..
5. Agrawal, R. Database Mining: A Performance Perspective [Text] / R. Agrawal, T. Imielinski, A. Swami // IEEE Transactions on Knowledge and Data Engineering. - 1993. - December. - P. 914-925.
6. Lippman, R.P. An introduction to computing with neural nets [Text] / R.P. Lippman // IEEE ASSP Magazine. - 1987. - P. 4-22.
7. Medsker, L.R. Recurrent neural networks: design and applications [Text] / L.R. Medsker, L.C. Jain // Boca Raton, FL: CRC Press. - 2000. - P. 25-27.
8. Elman, J.L. Finding structure in time Cognitive Science [Text] / J.L. Elman. - Vol. 14. - 1990. - P. 179-211.
9. Hyndman Another look at measures of forecast accurac [Text] / Hyndman, Rob J. Koehler, Anne B. - International Journal of Forecasting. P.679-688.
10. Jorrit Vander Mynsbrugge (2010). "Bidding Strategies Using Price Based Unit Commitment in a Deregulated Power Market", K.U. Leuven. 8
11. Каньовська, Д.В. Формування автоматизованої технології управління місцевою роботою на основі використання автономних модульних поїздів. Автореферат дисертації на здобуття наукового ступеня канд. техн. наук. Харків. УкрДАЗТ, 2013. - 21с.

Шандер О.Э. Формирование модели прогнозирования объемов вагонопотоков на станциях железнодорожного полигона. Работа посвящена вопросу разработки прогнозной модели, которая одновременно выдавала достаточно точную информацию о реальных объемах вагонопотоков на станциях и время предъявления вагонов на соответствующие станции. Сформирована архитектура искусственной нейронной сети Элмана, содержащая два входа и два выхода и является основой созданной прогнозной модели.

Ключевые слова: парк грузовых вагонов, железнодорожный полигон, вагонопотоки, модель прогнозирования, нейронные сети.

Shander O. The formation of car traffic volume prediction model at railway ground stations. The work is devoted to the development of a prediction model, which simultaneously produced sufficiently accurate information about real volumes of car traffic volumes at stations and the time of wagon presentation at appropriate stations. The architecture of an artificial Elman neural network comprising two inputs and two outputs which is the foundation of a created prediction model has been formed.

Key words: freight car fleet, railway ground, car traffic volumes, prediction model, neural networks.

Рецензент д.т.н., професор Огар О.М. (УкрДАЗТ)

Поступила 16.01.2015 г.