

ІНТЕЛЕКТУАЛЬНА КОМП'ЮТЕРНО-ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ПРОГНОЗУ ON-LINE ПРОДАЖІВ В ЕЛЕКТРОННИХ МАГАЗИНАХ

У статті розроблено формальні моделі, методи, інструменти та алгоритми, що у своїй сукупності формують інтелектуальну комп'ютерно-інформаційну технологію прогнозу on-line продажів в електронних магазинах Інтернету, яка дає змогу оптимізувати процеси прийняття управлінських рішень, логістики тощо. В основу даної інтелектуальної комп'ютерно-інформаційної технології покладено одну із концепцій нового тренду Data Science, а саме – концепцію Random Forest. Розроблена інтелектуальна комп'ютерно-інформаційна технологія може бути імплементована в комп'ютерно-інформаційні системи аналізу і прогнозу продажів.

Ключові слова: прогнозування продажів, дерева прийняття рішень, random forest.

A. MUSHAK

Ternopil National Economic University

INTELLIGENT COMPUTER-INFORMATION TECHNOLOGY FORECASTING ONLINE SALES IN E-SHOPS

Abstract – In the article formal models, methods, toolware and algorithms have developed. It forms intelligent computer-information technology prediction on-line sales in e-shops of Internet. It give possibility optimizes process decision making, logistics. The foundation of intelligent computer-information technology is new trend of Data Science, namely Random Forest. Intelligent computer-information technology may be insert in the computer-information system of analysis and prediction sales.

Keywords: prediction sales, decision trees, random forest.

Вступ

В останні роки широке використання інтелектуальних інформаційних технологій в найрізноманітніших задачах економіки спричинило бурхливий розвиток нового напрямку економічної кібернетики – Data Science. В основі технічної реалізації вирішення задач Data Science у сфері економічної кібернетики використовуються високопродуктивні комп'ютерні системи, на кшталт, суперкомп'ютери, GRID-системи тощо, які у своїй сукупності формують апаратне забезпечення хмарних комп'ютерно-інформаційних технологій.

У свою чергу розвиток напрямку Data Science в економічній кібернетичі відбувається у вигляді часткових розгалужень: Business Intelligence, Prediction Analytics, Data Mining, Statistics Data Processing, Big Data тощо. Серед наведених розгалужень широке практичне застосування набула Prediction Analytics – прогнозна аналітика. Вона використовується в основі інформаційних систем для оптимізації управлінських рішень, ризик менеджменту тощо. Серед великого сімейства практичних бізнес задач економічної кібернетики вагоме місце займають задачі прогнозу продажів в електронній комерції, наприклад, прогнозу on-line продажів в електронних магазинах Інтернету.

Складність вирішення такого класу задач обумовлена не лише труднощами у використанні хмарних комп'ютерно-інформаційних технологій, а передусім їх алгоритмічного вирішення. Відомі алгоритмічні рішення задач прогнозу on-line продажів в електронних магазинах базуються на таких концепціях та інформаційно-технологічних засобах.

Огляд існуючих відомостей

До базових концепцій алгоритмічних рішень задач прогнозу on-line продажів в електронних магазинах відносяться такі: нейромережева (класи нейронних мереж – нейронні мережі у вигляді багатопшарових перцептронів, нейронні мережі Кохонена, нейронні мережі Хемінга тощо), регресійна (багатофакторна лінійна, логістична та, загалом, нелінійна регресія), концепція на основі машин опорних векторів тощо.

Серед існуючих інформаційно-технологічних засобів, які дають змогу на вище згаданих концепціях розробляти інтелектуальні комп'ютерно-інформаційні технології вирішення задач прогнозу on-line продажів в електронних магазинах можна відзначити такі: Deductor [1], StatSoft [2], Matlab [3] тощо.

Разом з тим слід відзначити. Наведені концепції алгоритмічних рішень не завжди задовольняють апріорно заданим вимогам щодо точності прогнозу, оптимальності використання ресурсів хмарних комп'ютерно-інформаційних технологій. А існуючі інформаційно-технологічні засоби є універсальними, проте не завжди оптимальними щодо розробки інтелектуальних комп'ютерно-інформаційних технологій для вирішення задач прогнозу on-line продажів в електронних магазинах.

Тому у роботі розглядається можливість розробки нової інтелектуальної комп'ютерно-інформаційної технології для вирішення задачі прогнозу on-line продажів в електронних магазинах. Таким чином, метою даної роботи є наступне.

Мета

Розробити інтелектуальну комп'ютерно-інформаційну технологію для вирішення задач прогнозу on-line продажів в електронних магазинах Інтернету.

Постановка задачі

Розробити формальні моделі, інструменти та алгоритми, що у своїй сукупності формують інтелектуальну комп'ютерно-інформаційну технологію для вирішення задач прогнозу on-line продажів в електронних магазинах Інтернету.

Основна частина

Для розробки формальних моделей, інструментів та алгоритмів, що у своїй сукупності формуватимуть інтелектуальну комп'ютерно-інформаційну технологію для вирішення задач прогнозу on-line продажів в електронних магазинах Інтернету у роботі вибирається концепція Random Forest.

Для наочності висвітлення матеріалу у роботі розглядається одна абстрактна позиція (товар), яка наявна у продажі електронного магазину. Прогноз продажів по іншим позиціям здійснюватиметься аналогічно.

Нехай розглядуваним товаром є взуття, а кількість його пар – змінна y , що продається, залежить від трьох атрибутів: пори року – змінна x_1 , значення якої належить множині {зима, весна, літо, осінь}, розміру – змінна x_2 , значення якої належить множині {40, 42, 44} та кольору – змінна x_3 , значення якої належить множині {білий, чорний}.

У такому разі кількість пар взуття, що продається формально можна описати у вигляді функції від трьох змінних:

$$y = f(x_1, x_2, x_3).$$

На практиці кількість пар проданого взуття y залежить від більшої кількості факторів, тобто незалежних змінних: x_4, x_5, \dots . Але не завжди можна знати їх значення. Тому кількість проданих пар взуття y є випадковою величиною.

Враховуючи цей факт, задача прогнозу продажу кількості пар взуття y уточнюється наступним чином – потрібно здійснити прогноз продажу середньої кількості пар взуття, тобто $\mathbf{M}y$.

Нехай дано таблицю даних проданої кількості пар взуття попередньо, яка залежить від їх атрибутів (табл.1).

Очевидно, що завжди можна таку таблицю даних розбити на дві сукупності даних. Перша сукупність – навчаюча сукупність, друга – тестова сукупність.

Нехай перших п'ятнадцять елементів таблиці даних утворюють навчаючу сукупність, а других п'ятнадцять елементів таблиці даних утворюють тестову сукупність.

Виходячи з основної сутності концепції Random Forest потрібно згенерувати ліса дерев прийняття рішень різної висоти із випадковим генеруванням в них вершин і з випадковими навчаючими елементами з навчаючої сукупності. Значення проданої кількості пар взуття, змінної y розміщаються у листках дерев.

Ліс дерев прийняття рішень найменшої висоти – 1 представляє собою пеньки. Діючи за правилами, наведеними вище, одержуємо стохастичним чином ліс із пеньків (рис.1).

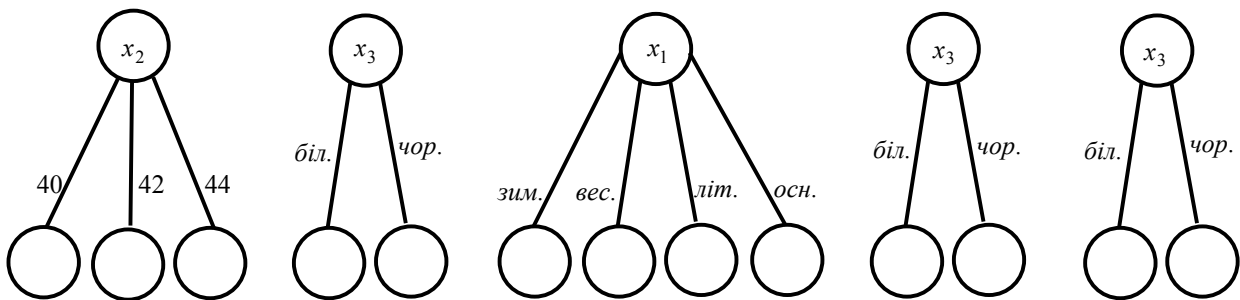


Рис.1. Випадковий ліс дерев прийняття рішень висотою 1

Як видно з рис.1 та табл.1, середнє значення проданої кількості пар взуття у першому зліва листку першого дерева дорівнює $\mathbf{M}y(40) = 80 \cdot 2/2 + 90 \cdot 0/2 \approx 80$, у другому листку – $\mathbf{M}y(42) = 80 \cdot 4/10 + 90 \cdot 6/10 \approx 86$, у третьому листку – $\mathbf{M}y(44) = 80 \cdot 1/3 + 90 \cdot 2/3 \approx 87$.

Відповідно до рис.1 та табл.1, середнє значення проданої кількості пар взуття у першому зліва листку другого дерева дорівнює $\mathbf{M}y(\text{біл.}) = 80 \cdot 4/9 + 90 \cdot 5/9 \approx 86$, у другому листку – $\mathbf{M}y(\text{чор.}) = 80 \cdot 3/6 + 90 \cdot 3/6 \approx 85$.

Відповідно до рис.1 та табл.1, середнє значення проданої кількості пар взуття у першому зліва листку третього дерева дорівнює $\mathbf{M}y(\text{зим.}) = 80 \cdot 1/2 + 90 \cdot 1/2 = 85$, у другому листку – $\mathbf{M}y(\text{вес.}) = 80 \cdot 0/3 + 90 \cdot 3/3 = 90$, у третьому листку – $\mathbf{M}y(\text{літ.}) = 80 \cdot 4/6 + 90 \cdot 2/6 \approx 83$, у четвертому листку – $\mathbf{M}y(\text{осн.}) = 80 \cdot 2/4 + 90 \cdot 2/4 = 85$.

Відповідно до рис.1 та табл.1 середні значення проданої кількості пар взуття у третьому та четвертому деревах аналогічні до другого дерева.

Таблиця 1

Кількість проданих пар взуття

	y	x_1	x_2	x_3
1.	80	зима	40	чорний
2.	80	осінь	42	чорний
3.	90	весна	44	білий
4.	80	осінь	42	чорний
5.	90	весна	42	білий
6.	80	літо	40	білий
7.	90	осінь	42	чорний
8.	90	зима	42	чорний
9.	80	літо	44	білий
10.	90	літо	42	білий
11.	90	весна	42	білий
12.	90	літо	44	білий
13.	80	літо	42	білий
14.	90	осінь	42	чорний
15.	80	літо	42	білий
16.	90	весна	42	білий
17.	90	літо	42	білий
18.	80	зима	40	чорний
19.	90	весна	42	білий
20.	90	літо	44	білий
21.	80	весна	42	чорний
22.	90	літо	40	чорний
23.	80	зима	44	білий
24.	90	весна	42	чорний
25.	80	літо	42	білий
26.	90	осінь	44	білий
27.	90	зима	44	білий
28.	90	осінь	42	чорний
29.	90	літо	42	білий
30.	80	весна	42	білий

Аналогічно до вище наведеного розглядається випадковий ліс з висотою дерев не більше ніж два яруси (рис.2).

З рис.2 та табл.1 видно, що середні значення проданої кількості пар взуття у першому зліва листку першого дерева дорівнює $M_y(\text{біл.}, 40) = 80 \cdot 1/1 + 90 \cdot 0/1 = 80$, у другому листку – $M_y(\text{біл.}, 42) = 80 \cdot 2/5 + 90 \cdot 3/5 = 86$, у третьому листку – $M_y(\text{біл.}, 44) = 80 \cdot 1/3 + 90 \cdot 2/3 \approx 87$, у четвертому листку – $M_y(\text{чор.}, 40) = 80 \cdot 1/1 + 90 \cdot 0/1 = 80$, у п'ятому – $M_y(\text{чор.}, 42) = 80 \cdot 2/5 + 90 \cdot 3/5 = 86$, у шостому – $M_y(\text{чор.}, 44) = 0$.

Відповідно до рис.2 та табл.1, середні значення проданої кількості пар взуття у першому зліва листку другого дерева дорівнює $M_y(\text{зим.}, \text{біл.}) = 0$, у другому листку – $M_y(\text{зим.}, \text{чор.}) = 80 \cdot 1/2 + 90 \cdot 1/2 = 85$, у третьому листку – $M_y(\text{вес.}, \text{біл.}) = 80 \cdot 0/3 + 90 \cdot 3/3 = 90$, у четвертому листку – $M_y(\text{вес.}, \text{чор.}) = 0$, у п'ятому – $M_y(\text{літ.}, \text{біл.}) = 80 \cdot 4/6 + 90 \cdot 2/6 \approx 83$, у шостому – $M_y(\text{літ.}, \text{чор.}) = 0$, у сьомому – $M_y(\text{осн.}, \text{біл.}) = 0$, у восьмому – $M_y(\text{осн.}, \text{чор.}) = 80 \cdot 2/4 + 90 \cdot 2/4 = 85$.

Також розглядається випадковий ліс з висотою дерев не більше ніж три яруси, приклад дерева якого наведено на рис.3.

Для даної задачі розглядається лише три види випадкових лісів. Загалом, кількість випадкових лісів відповідає кількості атрибутів від яких залежить цільова змінна, тобто та, яка прогнозується.

З рис.3 та табл.1 можна обчислити середні значення проданої кількості пар взуття у кожному із листку дерева прийняття рішень висотою у три яруси, тобто: $M_y(\text{зим.}, 40, \text{біл.})$, $M_y(\text{зим.}, 40, \text{чор.})$, $M_y(\text{зим.}, 42, \text{біл.})$, $M_y(\text{зим.}, 42, \text{чор.})$, $M_y(\text{зим.}, 44, \text{біл.})$, $M_y(\text{зим.}, 44, \text{чор.})$, $M_y(\text{вес.}, 40, \text{біл.})$, $M_y(\text{вес.}, 40, \text{чор.})$, $M_y(\text{вес.}, 42, \text{біл.})$, $M_y(\text{вес.}, 42, \text{чор.})$, $M_y(\text{вес.}, 44, \text{біл.})$, $M_y(\text{вес.}, 44, \text{чор.})$, $M_y(\text{літ.}, 40, \text{біл.})$, $M_y(\text{літ.}, 40, \text{чор.})$, $M_y(\text{літ.}, 42, \text{біл.})$, $M_y(\text{літ.}, 42, \text{чор.})$, $M_y(\text{літ.}, 44, \text{біл.})$, $M_y(\text{літ.}, 44, \text{чор.})$,

$\text{My}(\text{осн.}, 40, \text{біл.}), \quad \text{My}(\text{осн.}, 40, \text{чор.}), \quad \text{My}(\text{осн.}, 42, \text{біл.}), \quad \text{My}(\text{осн.}, 42, \text{чор.}), \quad \text{My}(\text{осн.}, 44, \text{біл.}),$
 $\text{My}(\text{осн.}, 44, \text{чор.}).$

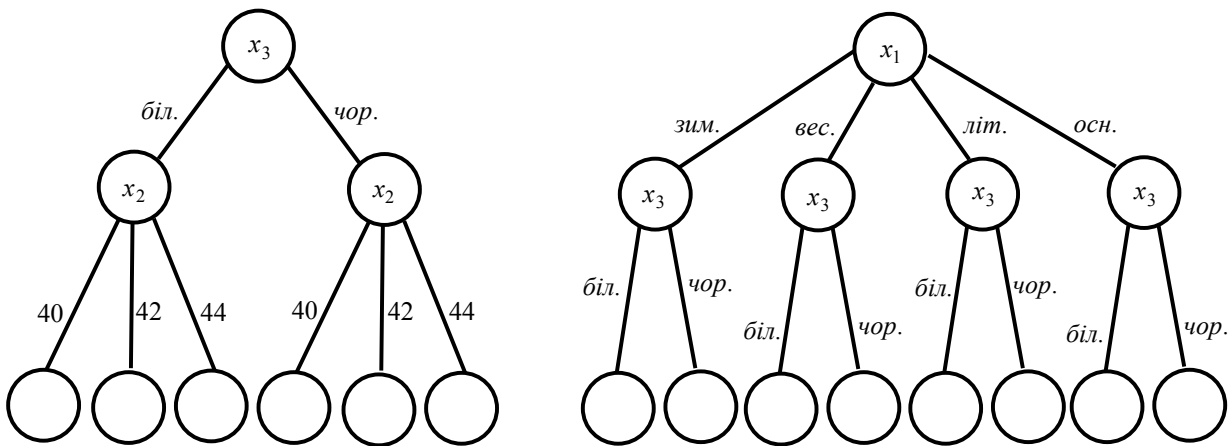


Рис.2. Випадковий ліс дерев прийняття рішень висотою 2

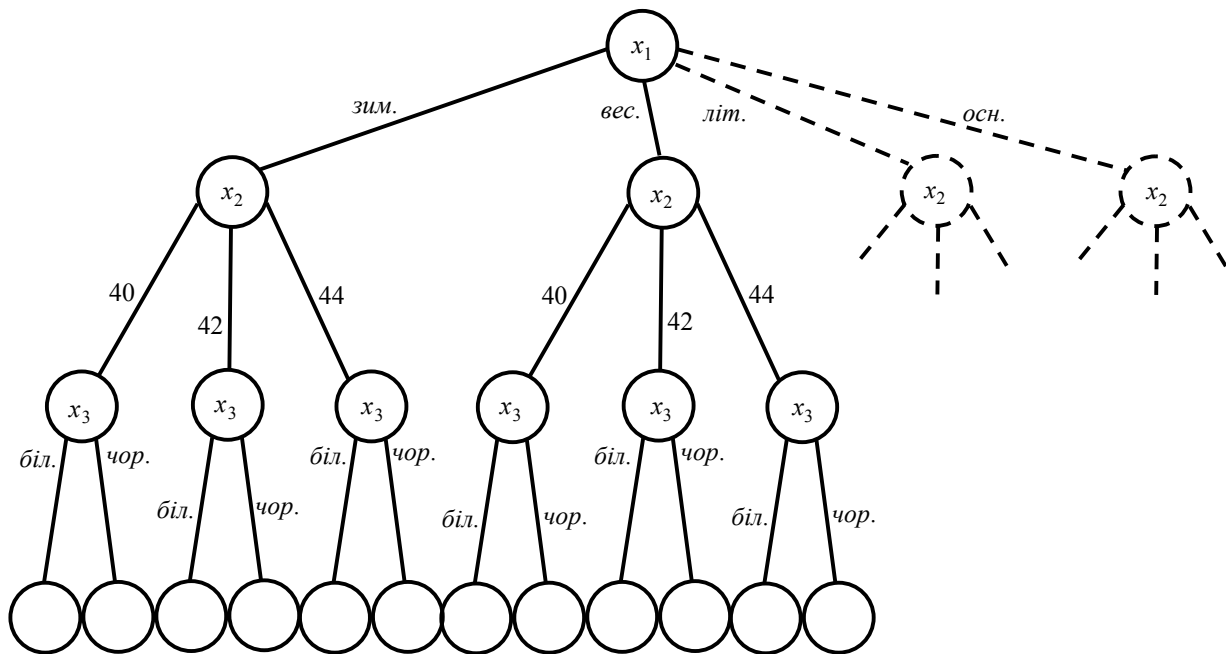


Рис.3. Приклад дерева прийняття рішень висотою 3 з випадкового лісу

Прийняття рішення щодо прогнозу значення продажу тої чи іншої кількості пар взуття здійснюється кожним окремим деревом прийняття рішень випадкового лісу. Загальне рішення випадкового лісу щодо прогнозного значення продажу кількості пар взуття одержують шляхом голосування усіх дерев прийняття рішень випадкового лісу. Остаточним вважається те рішення за яке проголосувало найбільше дерев прийняття рішень даного випадкового лісу. Також можливий випадок, коли виконується зважене голосування дерев прийняття рішень випадкового лісу. Вага кожного дерева прийняття рішень у голосуванні визначається його точністю прийнятого рішення, яка визначається на основі тестової сукупності.

Загалом потрібно відзначити, що кількість дерев прийняття рішень у кожному із випадкових лісів вибирається максимально можливою для забезпечення стійкості статистичних оцінок статистичних параметрів.

Інтегральне та остаточне рішення сукупності випадкових лісів щодо прогнозного значення продажу кількості пар взуття одержують шляхом їх голосування.

Таким чином, усе вище наведене: формальні моделі, методи, інструменти та алгоритми у своїй сукупності формують інтелектуальну комп'ютерно-інформаційну технологію прогнозу on-line продажів в електронних магазинах Інтернету.

Тестування даної інтелектуальної комп'ютерно-інформаційної технології прогнозу on-line продажів в електронних магазинах виконується на тестовій сукупності, тобто на других п'ятнадцяти елементах таблиці даних, які мають порядкові номери з 16 по 30.

Процедура тестування роботи технології виконується наступним чином. Кожен зокрема елемент

тестової сукупності тестує роботу технології, потім кожна пара елементів тестової сукупності тестує роботу технології, пізніше кожна трійка тестової сукупності тестує роботу технології, кожна четвірка, кожна п'ятірка і т.д. В результаті одержують множину значень похибки. Ця множина значень представляє собою вибірку випадкових значень похибки прогнозу. Тому для повного представлення похибки прогнозу обчислюється гістограма щільності розподілу імовірностей. У випадку, коли потрібно похибку представити у вигляді числового значення, то по вибірці її значень обчислюються статистичні оцінки: математичного сподівання, дисперсії, а також довірчі імовірності цих значень.

Висновки

У роботі розроблено формальні моделі, методи, інструменти та алгоритми, що у своїй сукупності формують інтелектуальну комп'ютерно-інформаційну технологію прогнозу on-line продажів в електронних магазинах Інтернету, яка дає змогу оптимізувати процеси прийняття управлінських рішень, логістики тощо.

Література

1. www.basegroup.ru.
2. www.statsoft.ru.
3. www.mathwork.com.

References

1. www.basegroup.ru.
2. www.statsoft.ru.
3. www.mathwork.com.

Рецензія/Peer review : 6.1.2015 р. Надрукована/Printed :26.1.2015 р.
Рецензент: д.т.н., проф. Пастух О.А.