УДК 537.331.322

С.В. Антоненко, Т.М. Булана, Б.В. Молодець

МОДЕЛЮВАННЯ ПРОГНОЗУВАННЯ НЕБЕЗПЕКИ ВИНИКНЕННЯ НАДЗВИЧАЙНИХ СИТУАЦІЙ

Анотація. Робота присвячена розробці моделі прогнозування пожеж. Поставлена задача полягає у прогнозуванні часових рядів за допомогою моделей AR, MA, ARIMA, проведенню аналізу отриманих результатів.
Ключові слова: прогнозування, пожежі, моделі ARIMA, MA, AR, часові ряди.

Глобальні процеси зміни клімату та стрімкий розвиток цивілізації призвели до значного підсилення кризових явищ як стихійної природи, так і технічного характеру. Світова спільнота виділяє багато уваги моніторингу надзвичайних ситуацій на усіх стадіях їх розвитку. Організація Об'єднаних Націй (ООН) навіть ввела поняття «цикл розвитку надзвичайної ситуації», в якому окреме місце відводиться етапу підготовки: аналізу ризику та ранньому попередженню катастрофи.

Бурний розвиток інформаційний технологій та лавиноподібне збільшення об'ємів допустимих даних різноманітної природи, зокрема супутникових, активізували дослідження в області геопросторового аналізу ризиків надзвичайних ситуацій. Оскільки геопросторовий аналіз ризиків припускають обробку великих об'ємів з різних джерел та побудови складних просторових моделей, рішення цієї задачі потребує створення відповідних автоматизованих інформаційних технологій на базі міжнародних стандартів обміну геопросторової інформації.

Дана робота буде цікава для державних структур, яка буде надавати можливість їм рівномірно розподілити сили МНС (Міністерство надзвичайних ситуацій) і буде актуальною для страхових компаній, які страхують нерухомість, щоб автоматизувати відображення можливого ризику.

При проведенні аналізу вхідних даних було вирішено спершу візуалізувати їх на карті та для кращого сприйняття відобразити штати з найбільшою кількістю пожеж яскраво червоного кольору, а з найменшою – темно-червоного кольору. В результаті ми отримали наступне (рис 1).

Рисунок 1 – Мапа, на якій відображена кількість пожеж за 23 роки

Виходячи з цього ми побудуємо наступну математичну модель: t_i– день з часового даного проміжку; x_i – пожежі, які відбулися в момент часу t_i в вибраному штаті.

Авторегресійні моделі широко використовуються для опису стаціонарних випадкових процесів. Характерною особливістю стаціонарних часових рядів є те, що їх імовірнісні властивості рядів не змінюються в часі. Інакше кажучи, функції розподілу стаціонарних динамічних рядів не змінюються при зсуві часу [1, 2].

Загальний вигляд моделі авторегресії p-ого порядку – AR (p) може бути виражений наступним рівнянням:

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_p y_{t-p} + \varepsilon_t \quad (t = 1, 2, \ldots, n)$$

де β_1, β_2 ,…,β_p – константи; ε_t – випадкові помилки.

AR(p) модель описує досліджуваний процес в момент t в залежності від його значень у попередні моменти t-1, t-2,…,t-p.

Модель передбачає майбутню поведінку, засновану на минулій поведінці. Вона використовується для прогнозування, коли є певна кореляція між значеннями в часі ряду і значеннями, які передують і процвітають в них [3].

В AR-моделі значення змінної результату (Y) в деякій точці t в часі є як «регулярна» лінійна регресія - безпосередньо пов'язана зі змінною (X). Там, де різні моделі лінійної регресії і AR відрізняються один від одного, Y залежить від X і попередніх значень для Y.

Процес AR являє собою приклад стохастичного процесу, який має ступінь невизначеності або випадковості. Зазвичай процес стає «досить близьким», оскільки він може бути корисний в більшості сценаріїв.

Моделі ковзного середнього (MA) представляють стаціонарний процес у вигляді лінійної комбінації послідовних значень білого шуму. Такі моделі виявляються корисними в якості самостійних описів стаціонарних процесів, а також в якості доповнення до моделей авторегресії для більш детального опису шумової складової.

Модель виражена наступним рівнянням:

$$y_t = \varepsilon_t - \gamma_1 \varepsilon_{t-1} - \gamma_2 \varepsilon_{t-2} - \dots - \gamma_q \varepsilon_{t-q}$$

де γ_1, $\gamma_2$2, ..., γ_q - параметри моделі.

З урахуванням властивостей помилки ε_t, можна побудувати автокореляційну функцію моделі MA (q). Її коефіцієнт коваріації q-ого порядку визначається наступним чином:

$$y_\tau = M(y_t, y_{t-\tau}) = M\left[(\varepsilon_t - \gamma_1\varepsilon_{t-1} - \gamma_2\varepsilon_{t-2} - \dots - \gamma_q\varepsilon_{t-q}) \times (\varepsilon_{t-\tau} - \gamma_1\varepsilon_{t-\tau-1} - \gamma_2\varepsilon_{t-\tau-2} - \dots - \gamma_q\varepsilon_{t-\tau-q})\right]$$

Авторегресійне інтегроване ковзне середнє (auto regressive integrated moving average, ARIMA) є узагальненням моделі авторегресійного ковзного середнього (ARMA). У загальному вигляді модель ARMA (p, q), де p - порядок авторегресії, q - порядок змінного середнього, виглядає наступним чином [3]:

$$y_t = \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \varepsilon_t + \theta_1 y_{t-1} + \dots + \theta_q y_{t-q}$$

Якщо процес виявляється нестаціонарним і для приведення його до стаціонарного виду треба було взяти кілька різниць, то модель стає ARIMA (p, d, q), де d - порядок різниці.

До очевидних переваг можна віднести те, що ці моделі мають дуже чітке математико-статистичне обґрунтування, що робить їх одними з найбільш науково обґрунтованих моделей з усієї безлічі моделей прогнозування тенденцій у тимчасових рядах. Також перевагою є формалізована і детально розроблена методика, слідуючи якій можна підібрати модель, найбільш підходящу до кожного конкретного часового ряду. Процедура перевірки моделі на адекватність досить проста, а розроблені методики з автоматичного підбору найкращої ARIMA полегшують роботу прогнозиста. Крім того, точкові і інтервальні прогнози випливають з самої моделі і не вимагають окремого оцінювання.

Один з явних недоліків моделей полягає у вимогах до рядів даних: для побудови адекватної моделі ARIMA потрібно не менше 40 спостережень. Дру-

 ISSN 1562-9945

гим серйозним недоліком є не адаптивність моделей авторегресії: при отриманні нових даних модель потрібно періодично переоцінювати, а іноді – і переідентифікувати.

Алгоритм роботи: аналіз даних, аналіз динаміки часового ряду;дослідження тенденцій часового ряду;перевірка на стаціонарність;оцінка моделей ARIMA; перевірка моделі на адекватність.

Побудова моделі AR і моделі MA передбачає стаціонарність часових рядів. Стаціонарний тимчасової ряд - це той, чиї властивості не залежать від часу, в яке спостерігається серія. Таким чином, тимчасові ряди з тенденціями або сезонністю не є стаціонарними - тренд і сезонність впливатимуть на значення часових рядів при різних раз. випадку специфіки даних, отриманих для аналізу, отримуємо, що вони не стаціонарні, бо в них присутня сезонність. Це побачимо на рис. 2.

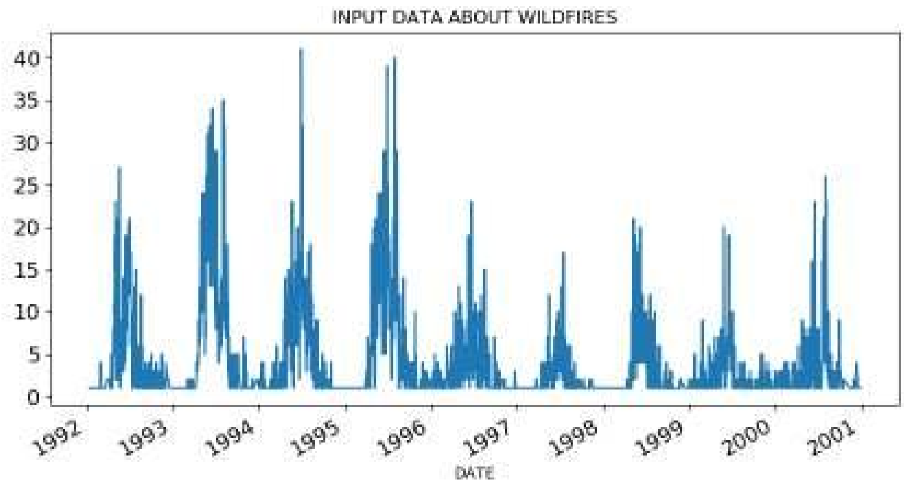

Рисунок 2 - Не стаціонарний часовий ряд

Також є аналітичний тест, що перевіряє часовий ряд на стаціонарність – тест Дікі - Фуллера. Часовий ряд має одиничний корінь або порядок інтеграції один, якщо його перші різниці утворюють стаціонарний ряд. У термінах математики ця умова записується як:

$\{y(k)\} \sim I(1)$ якщо $\Delta y(k) = y(k) - y(k-1)$ стаціонарний ряд $\{\Delta y(k)\} \sim I(0)$

За допомогою цього тесту визначають значення коефіцієнта в авторегресіоному рівнянні першого порядку AP (1):

$$y(k) = p \cdot y(k-1) + \varepsilon(k)$$

$y(k) -$ часовий ряд, $\varepsilon(k) -$ похибка.

Якщо $p=1$ то процес має одиничний корінь, в цьому випадку y не стаціонарний. Степінь інтегрованості процесу дорівнює $I(1)$. то

Якщо $0<p<1$, то ряд стаціонарний та має степінь інтегрованості $I(0)$.

Наведене авторегресійне рівняння можна переписати у вигляді:

$\Delta y(k)=b\cdot y(k-1)+e(k)$, де $b=p-1$, а $\Delta-$оператор різниці першого порядку $\Delta y(k)=y(k)-y(k-1)$.

Збільшена статистика Дікі - Фуллера, використовувана в тесті ADF, є негативним числом, і чим більше негативним вона є, тим сильніше відхилення гіпотези про те, що існує одиничний корінь.

Звичайно, це тільки на деякому рівні впевненості. Тобто, якщо статистика тесту ADF позитивна, можна автоматично прийняти рішення не відхиляти нульову гіпотезу одиничного кореня. В одному прикладі, з трьома затримками, значення -7.287006e+00 становило відхилення при p-значення - 7.287006e+00 (рис. 3).

Results of Dickey-Fuller Test:	
Test Statistic	-7.287006e+00
p-value	1.450008e-10
#Lags Used	1.500000e+01
Number of Observations Used	4.761000e+03
Critical Value (1%)	-3.431724e+00
Critical Value (5%)	-2.862147e+00
Critical Value (10%)	-2.567093e+00

Рисунок 3 - Результати розширеного тесту Дікі - Фуллера на данних про пожежі в США за 1992-2001

Деякі випадки можуть ввести в оману – часовий ряд з циклічним поведінкою (але не трендом або сезонністю) є стаціонарним. Це тому, що цикли не мають фіксованої довжини, тому, перш ніж ми спостерігаємо серію, ми не можемо бути впевнені, де будуть піки і западини циклів.

Один із способів зробити тимчасової ряд стаціонарним – обчислити різницю між послідовними спостереженнями. Це відомо як відмінність.

Перетворення, такі як логарифми, можуть допомогти стабілізувати дисперсію тимчасового ряду. Різниця може допомогти стабілізувати середнє тимчасового ряду шляхом усунення змін рівня часових рядів і, таким чином, усунення тенденції і сезонності.

Висновки. В даній роботі було проведено порівняльний аналіз підходів прогнозування виникнення пожеж. Зокрема були реалізовані математичні моделі прогнозування виникнення пожежі та був розроблений продукт автоматизованого прогнозування ризику виникнення пожеж, було досліджено декілька підходів тестування часового ряду: візуальний метод, розширений метод Дікі - Фуллера. Для прогнозування ризику виникнення пожежі були використані наступні моделі: AR, MA, ARIMA. В результаті тестування було виявлено, що модель ARIMA дає найкращі результати.

ЛІТЕРАТУРА

1. Кендэл М. Временные ряды / Кендэл М. – М.: Финансы и статистика, 1981. – 202 с.

2. Рунова Л.П., Рунов И.Л. Анализ временних рядов и прогнозирование. Учебно-методические материалы по дисциплине Методы социально-экономического прогнозирования для студентов со специальности Математические методы в экономике / Л.П. Рунова, И.Л. Рунов – Ростов-на-Дону, РГУ, 2006. – 97 с.

3. Бокс Дж. Анализ временних рядов прогноз и управление / Дж. Бокс, Г. Дженкинс [Под ред. В.Ф. Писаренко]. — М.: Мир, 1974. — 406 с.

REFERENCES

1. Kendel M. Vremennyie ryadyi / Kendel M. – M.: Finansyi i statistika, 1981. – 202 s.

2. Runova L.P., Runov I.L. Analiz vremennih ryadov i prognozirovanie. Uchebno-metodicheskie materialyi po distsipline Metodyi sotsialno-ekonomicheskogo prognozirovaniya dlya studentov co spetsialnosti Matematicheskie metodyi v ekonomike / L.P. Runova, I.L. Runov – Rostov-na-Donu, RGU, 2006. ¬– 97 s.

3. Boks Dzh. Analiz vremennih ryadov prognoz i upravlenie / Dzh. Boks, G. Dzhenkins [Pod red. V.F. Pisarenko]. — M.: Mir, 1974. — 406 s.