УДК 004.896.8

МАТЕМАТИЧНЕ ТА ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ ДЛЯ ЗНАХОДЖЕННЯ АНОМАЛІЙ В ТЕХНОЛОГІЧНИХ ДАНИХ

Поворозник Б.С.[1)],**Марценюк Є.О.**[2)], **Куривчак Д.І.**[3)], **Мархоцький Н.Ю.**[4)], **Даньків А.В.**[5)]
Західноукраїнський національний університет
[1)]магістрант;[2]к.т.н., доцент;[3-4)]магістрант;[5)]аспірант

I. Постановка проблеми

Розуміння та використання даних, які генеруються технологічними системами, стає все важливішим для ефективного управління та забезпечення надійності виробничих процесів.

Однією з ключових переваг кіберфізичних систем є їх здатність інтегрувати фізичні об'єкти з програмним забезпеченням, що дозволяє збирати, аналізувати та діяти на основі зібраних даних в реальному чи близькому до реального часу. Прогрес у цій області може значно підвищити ефективність, надійність та безпеку виробничих процесів [1].

Щодо складнощів у впровадженні інтелектуальних систем аналізу даних у виробничому середовищі, то згадані проблеми є цілком актуальними. Врахування нестаціонарності, складних залежностей та унікальної природи технологічних даних вимагає розробки високоефективних алгоритмів аналізу та моделей, які здатні працювати в реальному часі та реагувати на зміни виробничих умов .

Проте, розвиток науки про дані та швидкий прогрес в області штучного інтелекту та машинного навчання відкривають нові можливості для подальшого розвитку і вдосконалення систем виявлення аномалій у виробничих процесах. Із поглибленням розуміння цих складних проблем та застосуванням передових технологій, можна досягти значних досягнень у покращенні безпеки та продуктивності виробничих систем.

II. Мета роботи

Метою дослідження є розробка математичного та програмного забезпечення для знаходження аномалій в технологічних даних.

III. Метод виявлення аномалій в технологічних даних

Складність виявлення аномалій у технологічних сигналах полягає у їхній "нестаціонарності". При цьому з урахуванням постановки завдання виявлення аномалій повинно відбуватися без вчителя, тобто без будь-якої апріорної інформації про властивості сигналу. У той же час, математичний апарат системи повинен мати інваріантні властивості аналізованих даних. Ці фактори вимагають від алгоритму виявлення аномалій адаптивних властивостей, що дозволяють витягувати інформативні ознаки із сигналів без додаткової інформації.

Тому пропонується підхід до побудови алгоритму G з використанням перетворення Гільберта для отримання амплітудно-частотного спектра та подальшого його аналізу. Основна ідея пропонованого підходу полягає у розкладанні досліджуваного сигналу $X(t)$ в амплітудний спектр та застосуванні моделі виявлення M аномалій до кожного рівня розкладання.

При цьому модель виявлення аномалій як вхідну інформацію приймає миттєву амплітуду, одержану шляхом застосування перетворення Гільберта до внутрішніх мод сигналу (IMF), отриманих в результаті емпіричної модової декомпозиції сигналу (метод Хуанга). Зауважимо, що для аналізу використовується саме миттєва амплітуда для відповідної IMF, а миттєві частота і фаза не беруться до уваги через низьку локалізацію особливостей сигналу. Результати виявлених аномалій кожного рівня розкладання представляються як підсумковий вектор міток. Нижче представлений формальний покроковий опис алгоритму:

Крок 1. Згідно з постановкою задачі сигнал $X(t)$ поділяється на навчальну $X(t)^{train}$ та тестову вибірку $X(t)^{test}$.

Крок 2. Для кожної частини сигналу виконується процедура вилучення $IMF(t)^{train}$ та $IMF(t)^{test}$ таким чином, що:

$$X(t) = \sum_{j=1}^{n} IMF_j(t) + r_n(t) \quad (1)$$

де n - кількість емпіричних мод, $r_j = r_{j-1} - IMF_j(t)$.

Крок 3. Виявляється мінімальний рівень $T_{\min}$ розкладання IMF через порівняння числа рівнів розкладання для $IMF(t)^{train}$ та $IMF(t)^{test}$:

$$T_{\min} = \min(T^{train}, T^{test}) \tag{2}$$

Крок 4. Для $IMF(t)^{train}$ і $IMF(t)^{test}$ застосовується перетворення Гільберта, після чого обчислюються образи Гільберта $H(t)^{train}$ та $H(t)^{test}$:

$$H(t)^{train} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{IMF(\tau)^{train}}{t-\tau} d\tau \tag{3}$$

$$H(t)^{test} = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{IMF(\tau)^{test}}{t-\tau} d\tau \tag{4}$$

де інтеграли в (3,4) розуміються у сенсі головного (P.V. – principal value) значення Коші.

Крок 5. Модель M навчається на миттєвій амплітуді $\left|H(t)_k^{train}\right|$, де $k = 1..T_{\min}$:

$$M(\varphi, \theta, |H(t)_k^{train}|) \rightarrow M' \tag{5}$$

де M' – навчена модель.

Крок 6. Далі для $H(t)_k^{test}$ виконується процедура виявлення аномалій та формується вектор прогнозу F:

$$M'(\varphi, \theta, |H(t)_k^{test}|) \rightarrow F \tag{6}$$

На рисунку 1 представлено графічне, деталізоване представлення запропонованого методу. Підхід складається із двох блоків. Перший блок виконує процедуру перетворення Гільберта для аналізованого сигналу, другий блок навчає та тестує моделі для кожного отриманого образу Гільберта $H(t)$. При цьому як моделі виявлення можуть застосовуватись різні моделі M для кожного сигналу. Далі прогноз кожної моделі об'єднуються у спільний вектор F оцінок аномалій. Відмінною рисою даного підходу є те, що кожен елемент спектра сигналу аналізується окремою моделлю M_k. У разі виникнення аномалії в сигналі, на певному рівні розкладання, дана аномалія локалізується вираженим шляхом і буде виявлена відповідною моделлю. При цьому рівень розкладання контролюється автоматично за рахунок обчислення $T_{\min}$, що дозволяє аналізувати максимальну кількість рівнів розкладання. Це дозволяє підвищити як точність виявлення, а й інтерпретованість підходу.

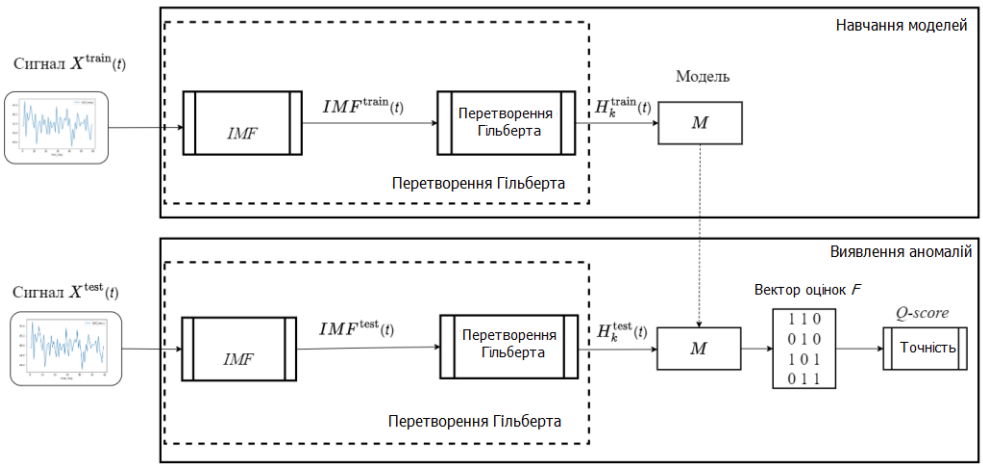

Рисунок 1– Схема методу виявлення аномалій

Висновок

Запропонований підхід до виявлення аномалій на основі перетворення Гільберта відкриває нові можливості для ефективного аналізу сигналів і виявлення незвичайних подій. Використання цього методу дозволяє локалізувати аномалії в елементах спектру сигналу та аналізувати їх окремо, що сприяє більш точному виявленню та ідентифікації. Основна новизна цього підходу полягає у тому, що кожен елемент амплітудного спектра сигналу аналізується окремо моделлю виявлення. Це означає, що кожен компонент сигналу розглядається індивідуально, що дозволяє ефективно виявляти аномалії, які можуть відбуватися на різних частотах або в різний час.

Список використаних джерел

1. В.П. Шкодирєв, К.І. Ягафаров, В.А. Баштовенко, Є.Е. Ільїна. Огляд методів виявлення аномалій в потоках даних. URL: http://ceur-ws.org/Vol-1864/paper_33.pdf