

## КЛАСТЕРНИЙ АНАЛІЗ ДЛЯ ЗМЕНШЕННЯ НАВАНТАЖЕННЯ В МЕРЕЖАХ ІОТ

**Гребініченко М.В. Бугаєнко Ю.М.**

*Інститут телекомунікаційних систем КПІ ім. Ігоря Сікорського, Україна*

*Email: grebinichenko.maria@gmail.com*

### **Cluster analysis for reducing load in IoT Networks**

In this article is studying problem of proceeding big amount of data in IoT networks. One of the methods for resolving this problem is cluster analysis. The benefits of application of this method are presented and described.

Зі стрімким розвитком розумних систем Internet Of Things, які мають обробляти великі масиви даних, з'явилася необхідність оптимізації традиційної архітектури мережі для забезпечення передачі та обробки цих даних з якомога більшою ефективністю та використовуючи при цьому меншу полосу пропускання.

Інтернет Речей можна символічно представити у вигляді такої формули:

*Інтернет Речей = Сенсори + Дані + Мережі + Послуги*

У будь-якій системі Інтернету Речей чільне місце займає процес передачі даних від сенсорів, що збирають інформацію, до серверів, що здійснюють її подальшу обробку та зберігання. За короткий проміжок часу по мережі може передаватись величезна кількість даних, що викликає перевантаження, а відповідно й затримки, втрати інформації та інші проблемні ситуації.

Класичного статистичного аналізу даних вже не достатньо при роботі з великими обсягами даних. Необхідний механізм більш оптимального аналізу даних та відповідно потрібною є модифікація архітектури мережі Інтернету Речей для забезпечення більш ефективної роботи з даними, а саме блоків передачі та обробки даних.

Блок обробки даних повинен мати невисоку обчислювану складність, і при тому дозволяти розподіляти їх. Одночасно блок передачі даних має передавати ці дані до інших компонентів системи згідно певних шаблонів, що будуть відповідати ефективній передачі саме цих даних, що збираються в мережі. Для різних наборів даних мають застосовуватись різні шаблони, наприклад, окремо для аудіо-, відеотрафіку і т.д. З цього випливає, що застосування цих шаблонів вимагає класифікації трафіку.

Ця задача має вирішуватись у блоці обробки даних, який ми також модифікуємо. Всередині нього повинні застосовуватись деякі логічні правила аби уникнути великих обсягів обчислень та класифікувати інформацію. Для формування цих правил ми будемо використовувати методи модифікованого алгоритму кластеризації та математичного апарату нечіткої логіки. Застосування підходу кластерного аналізу в системах Інтернету речей дозволить оптимально структурувати необхідні дані для передачі по мережі, і, відповідно, знизити

навантаження на неї.

**Метою кластерного аналізу** є призначення об'єктів в наборі даних у відповідні класи таким чином, щоб об'єкти в одному класі були більше схожими один на одного, ніж з тими, що в інших класах. Метод кластеризації відноситься до засобів, за допомогою яких формуються кластери або групи. Кластерний аналіз буде відноситись до загальної послідовності кроків, які представляють аналіз.

Вибір методу кластеризації є критичним етапом процесу кластеризації. Під час вибору методу слід враховувати, що метод повинен бути спроектований для відновлення типів кластерів, які імовірно присутні в даних, бути ефективним при відновленні структур, для яких він був розроблений та мати можливість протистояти наявності помилок у даних.

В літературі запропоновано багато методів кластеризації. Було проведено дослідження по порівнянню методів кластеризації з точки зору здатності відновлювати кластерні структури. Для цих цілей підходять алгоритми нечіткої логіки, що мають одразу кілька переваг для використання у IoT-системах: можливість кластерного аналізу для наборів інформації, що вимірюються багатьма параметрами, забезпечення більшої гнучкості у розподілі даних на кластери.

Основна ідея нечіткої логіки полягає в тому, що інтелектуальний спосіб міркувань, що спирається на природну мову спілкування людини, не може бути описаний традиційними математичними формулами.

Вхідною інформацією для кластеризації служить таблиця спостережень  $T = \{t_1, t_2, \dots, t_M\}$ , де  $k$ -й елемент  $t_k = (x_k, y_k)$ ,  $x_k = [x_{k1}, x_{k2}, \dots, x_{kn}]$  - вектор вхідних значень,  $y_k$  — вихідне значення,  $M$  — число спостережень,  $x_k \in DX$ ,  $y_k \in DY$ . Нехай  $P(T)$  — множина всіх підмножин множини  $T$ . Нечітке  $c$ -розбиття  $T$  є множина  $\{A_i | A_i \in P(T), 1 \leq i \leq c, 2 \leq c \leq M-1\}$ ,  $c$  — число кластерів. Кластери характеризуються матрицею нечіткого розбиття  $F = [\mu_{ki}]$ , яка має такі властивості:

$$\forall_{k=1, M} \left( \sum_{i=1}^c \mu_{ki} = 1 \right), \quad \forall_{i=1, c} \left( 0 < \sum_{k=1}^M \mu_{ki} < M \right)$$

Центральне місце в нечіткій логіці займає нечіткий логічний висновок, результатом якого є чітке значення змінної. Для того щоб виконати нечіткий логічний висновок необхідні наступні умови:

1. Повинно існувати як мінімум одне правило для кожного терма вхідної змінної.
2. Для будь-якого терма вхідної змінної повинно бути хоча б одне правило, в якій цей терм використовується в якості передумови.
3. Між правилами не повинно бути суперечностей і кореляції.

На рис. 1 зображена послідовність дій при використанні процесу нечіткого логічного висновку.

На етапі фазифікації у блоці обробки інформації має визначатися шаблон, згідно якого у подальшому дані мають передаватися, та застосовуватись первний протокол щодо передачі даних.

Якщо розглянути приклад невеликої системи IoT, що одночасно збирає

інформацію з сенсорів температури, смартфона користувача, що відправляє певні дані зі свого смартфона до пристрою, підключеного до мережі, та приладу, що посилає сигнал аварійної ситуації. Цю інформацію можна розбити на кластери за типом трафіку, що передається, та за важливістю.



Рис. 1. Послідовність дій при використанні процесу нечіткого логічного висновку.

### **Переваги:**

1. Зменшення кількості та тривалості затримок при передачі інформації, часу на передачу інформації;
2. Зменшена кількість відмов, ймовірності втрати даних;
3. Зрозумілість та інтерпретування моделей;
4. Індивідуальний підхід передачі для кожного типу кластеризованих даних.

### **Недоліки:**

1. Високі обчислювальні витрати проведення нечіткого виведення;
2. Збільшення часу на обробку інформації.

**Висновки:** В процесі роботи над статтею було проаналізовано такий спосіб оптимізації навантаження в мережах як кластерний аналіз та застосування нечіткої логіки. У процесі досліджень ми дійшли висновку, що керування цими процесами може відбуватися на основі логічних правил. Було розглянуто певні методи кластерного аналізу, що дозволяють будувати такі правила та визначати приналежність даних до одного або іншого типу з подальшим застосуванням шаблонів передачі. Це створює передумови зменшення навантаження на мережу, обсягу трафіку. У подальшій роботі необхідно дослідити та запропонувати модифікацію блоків обробки та передачі даних мережі Інтернету Речей.

### **Література**

1. Lea R. Smart cities: An iot-centric approach / R. Lea, M. Blackstock. // Proceedings of the 2014 International Workshop on Web Intelligence and Smart Sensing, IWWISS'14. – 2014. – P. 12–15.
2. Almasri M. Data Fusion Models in WSNs: Comparison and Analysis / M. Almasri // Conference of the American Society for Engineering Education IEEE 2014.
3. Globa L.S. INTELLIGENT SUPPORT SYSTEM FOR E-HEALTH / L.S. Globa, I.O. Ishchenko, A.G. Zakharchuk // 20th International Multi-Conference on Advanced Computer Systems Międzynarodowe, Poland – 2016. – P. 8.
4. Куташов П. Д. Городские координатные АТС типа АТСК / П. Д. Куташов, Б. С. Лившиц, А. Л. Пошерстник, Г. Б. Ханін // М.: Связь – 1970. – С. 304.