

Артемчук В.О.

*к.т.н., с.н.с., Інститут проблем моделювання в енергетиці
ім. Г.Є. Пухова НАН України, м. Київ
ak24avo@gmail.com*

Яцишин А.В.

*д.т.н., с.н.с. Інститут проблем моделювання в енергетиці
ім. Г.Є. Пухова НАН України, м. Київ*

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ МОНІТОРИНГУ СТАНУ АТМОСФЕРНОГО ПОВІТРЯ

Аналіз світового досвіду свідчить про ефективність та перспективність сенсорних мереж як аналізаторів якості повітряного середовища. В умовах міст України така система моніторингу стану атмосферного повітря (МСАП) допоможе вирішити проблеми, що склалися у цій галузі, покращити технічне оснащення мережі та підвищити її оперативність в рамках зменшення техногенного впливу об'єктів енергетики на довкілля. Проте обов'язковою складовою моніторингу, окрім власне пунктів спостереження, є засоби аналізу даних, в т.ч. інтелектуального, на основі результатів яких відбувається управління екологічною безпекою. Не дивлячись на певне число робіт, в яких розглядаються питання інтелектуального аналізу даних мережі моніторингу стану атмосферного повітря, можна констатувати, що комплексно ці питання з врахуванням вимог та рекомендацій сучасного міжнародного та Українського законодавства досить докладно не розглядалися. Отже, розробка засобів інтелектуального аналізу даних мережі моніторингу стану атмосферного повітря в рамках зменшення техногенного впливу об'єктів енергетики на довкілля є актуальною науково-технічною проблемою, що потребує вирішення.

На основі публікацій [**Ошибка! Источник ссылки не найден.** та ін.] виявлено найбільш широко використовувані алгоритми інтелектуального аналізу даних – тобто алгоритмів виявлення прихованих закономірностей або взаємозв'язків між змінними у великих масивах необроблених даних: C4.5 створює класифікатор у вигляді дерева рішень (ДР). Для цього C4.5 надається набір даних, який представляє собою вже класифіковані речі. K-means створює k груп з набору об'єктів таким чином, щоб члени цієї групи були якомога більше схожі. Це популярний метод кластерного аналізу для вивчення набору даних. Нагадаємо, що кластерний аналіз – це сімейство алгоритмів, призначених для формування груп, де члени цих груп схожі один на одного сильніше, ніж на тих, хто в цій групі не перебуває (в даному контексті кластери та групи є синонімами). Метод опорних векторів (SVM) знаходить гіперплощину для класифікації даних в два класи. На відміну від C4.5 не використовує ДР. Алгоритм Apriori шукає асоціативні правила і застосовує їх до бази даних, що містить велику кількість транзакцій. Нагадаємо, що пошук асоціативних правил – це метод отримання даних для вивчення кореляцій і взаємозв'язку між змінними в базі даних. EM у добуванні даних найчастіше використовується як алгоритм кластеризації (як k-means) для виявлення знань. PageRank є алгоритмом посилального ранжирування для визначення відносної «важливості» будь-якого об'єкта в мережі об'єктів. AdaBoost - алгоритм посилення класифікаторів. kNN (k найближчих сусідів) - алгоритм для класифікації, що відрізняється від раніше описаних тим, що він є «лінивим учнем». Наївний баєсів класифікатор є не єдиним алгоритмом, а сімейством алгоритмів класифікації, які поділяють одне спільне припущення. CART означає дерево класифікації і регресії (Classification and Regression Tree). Це метод навчання способом побудови ДР, який видає або дерева класифікації, або регресії.

Ці та ряд інших алгоритмів ввійдуть до системи, основу структури (див. рис. 1) якої взято з [2]. Розробка такої системи дозволить вирішити ряд задач управління екологічною безпекою, зокрема в контексті зменшення техногенного впливу об'єктів енергетики на довкілля.

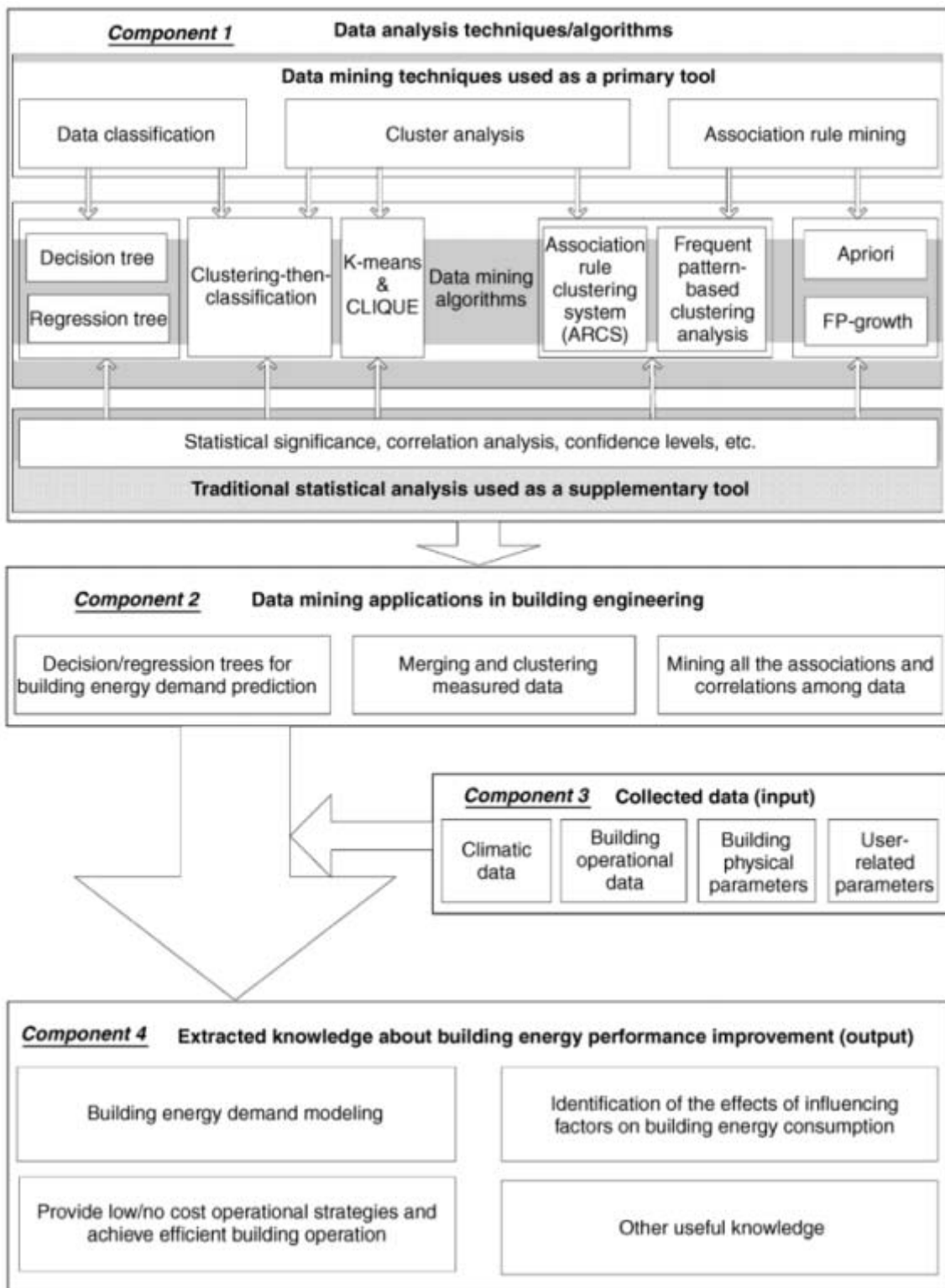


Рис. 1. Структура системи інтелектуального аналізу [2]

Список використаних джерел

1. Алгоритмы интеллектуального анализа данных / – Дата доступу 20.03.2018. – Режим доступу: <https://tproger.ru/translations/top-10-data-mining-algorithms/>
2. Yu, Z., Fung, B.C.M. and Haghghat, F. 2013. Extracting knowledge from building-related data – A data mining framework. BUILD SIMUL 6: 207-222.