

Методи кластеризації даних у технологічних процесах поліграфічного виробництва

Я.О. Меденець¹

Anotation – The main directions of perfection of printing processes automation have been offered. The parameter's clusterization description conducted in printing process. Cluster analysis foundations have been offered based on k-means method.

Ключові слова – друкарська рулонна ротатійна машина, кластеризація, параметри, друкарський процес.

Основними напрямками вдосконалення автоматизації друкарських процесів є: розробка комплексних систем управління агрегатами друкарської машини і параметрами друкарського процесу; розробка систем автоматизованого налаштування фарбового апарату; створення систем збору і відображення параметрів друкарського процесу. Зі всього різноманіття друкарських машин найбільш складні і продуктивні — рулонні ротатійні друкарські машини (PPM).

Під час друкарського процесу на основну консоль панелі управління багатофарбової PPM надходять параметри від різноманітних давачів: параметри швидкості руху полотна, натягу, зволоження, подачі фарби, суміщення кольорів, подібні. На якість друку також впливають й інші параметри - другорядні: якість фарби (відповідає за параметри насичення, яскравості) та зволожуючого розчину, температура в приміщенні, значення відносної вологості, граматура паперу та ін.

У деяких виробничих процесах кількість параметрів може досягати тисяч, а частота, з якими вони оновлюються, може доходити до 10 ГГц. Таким чином, для ефективного опрацювання великої кількості різноманітних даних, виникає необхідність кластеризації параметрів та декомпозиції оброблюваних даних.

Основні етапи функціонування інформаційної технології обробки кластеризованих даних показані на рис.1. Базові параметри потрапляють на модуль завантаження параметрів роботи, де порівнюються з біжучими параметрами. Завантаження відбувається усіма паралельними процесами одночасно. Відбувається обчислення та перевірка, чи дані параметри підпадають під прийнятні критерії.

У модулі представлення результатів на основі опрацювання усієї інформації отримуємо загальний результат, що показує повну характеристику виробничого процесу. Дані виводяться на екран панелі управління та фіксуються.

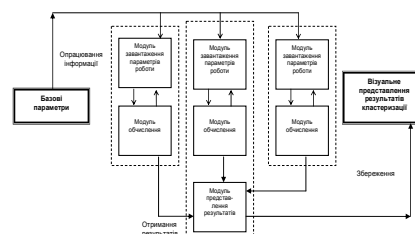


Рис. 1. Основні модулі інформаційної технології кластеризації технологічного процесу

На прикладі PPM можна відобразити одним елементом (центроїдом) наступні параметри, що формують кластери даних: швидкість машини – відноситься до кластеру швидкості; вузли машини (секції подачі рулонів паперу, друкарські пари, фальцапарат) – кластер вузлів; якість фарби (насичення, яскравість) – кластер якості відбитку...

Основна ідея кластеризації – виділити з елементів основні, а також центр кластера, який повинен задовільняти обмеженням і вимогам технологічного процесу. Опишемо параметри друкарської машини такими функціональними блоками:

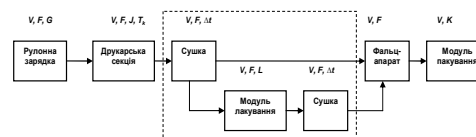


Рис. 2. Функціональна схема багатофарбової друкарської машини.

Продуктивність друкарської описуємо наступною залежністю:

$$Pr = < \forall V_i, F_i, G, J_p, \Delta t, L_p, K_k > \quad (1)$$

V – швидкість паперового полотна,

F – натяг стрічки,

$i = 1..N$, N – кількість фарбових апаратів,

G – якість паперу (граматура), g – ширина рулону,

J – якість фарби (яскравість, насиченість, густина),

$j = 1..S$, S – кількість фарб,

T_k – точність роботи,

$k = 1..k$, k – точність роботи окремих елементів,

Δt – зміна температури на ділянках сушки,

L_l – якість лаку,

$l = 0..1$, l – лакування,

K_m – блок пакування, m – задається оператором.

Завдання кластеризації найчастіше вирішується методами математичної статистики. Формальна постановка задачі кластеризації: нехай X – множина об'єктів, Y – множина номерів кластерів.

¹ Українська академія друкарства, вул. Підвальна, 17, Львів, 79000, УКРАЇНА, E-mail: lightsunrise@gmail.com

Задана функція відстані між об'єктами $p(x, x')$. Маємо кінцеву вибірку об'єктів:

$$X^m = \{x_1, \dots, x_m\} \subset X \quad (2)$$

Потрібно розбити вибірку на підмножини (кластери), що перетинаються, так, щоб кожен кластер складався з об'єктів, близьких до p , а об'єкти різних кластерів істотно відрізнялись. Кожному об'єкту $x_i \in X^m$ приписується номер кластера y_i .

Алгоритм кластеризації – функція $a: X \rightarrow Y$, що будь-якому об'єкту $x \in X$ ставить y відповідність номер кластера $y \in Y$. Об'єкти описуються p числовими ознаками $f_1(x), \dots, f_n(x), X = \mathbb{R}^n$. Кожен кластер $y \in Y$ описується p -мірною гаусівською щільністю $p_y(x) = N(x; \mu_y, \sum y)$ з центром $\mu_y = (\mu_{y1}, \dots, \mu_{yn})$ і діагональною коваріаційною матрицею $\sum y = \text{diag}(\sigma_{y1}^2, \dots, \sigma_{yn}^2)$. При цих уявленнях задача кластеризації співпадає з задачею розділення імовірностей розподілень і для її вирішення можна застосувати EM-алгоритм. На E-кроці, згідно формули Байеса, вираховують приховані змінні g_{iy} . Значення g_{iy} рівне імовірності того, що об'єкт $x_i \in X^m$ належить кластеру $y \in Y$. На M-кроці уточнюються параметри кожного кластеру $\mu_y, \sum y$, при цьому використовуються приховані змінні g_{iy} .

Найчастіше з практичною метою використовується метод **k-середніх**. Головною відмінністю є те, що у EM-алгоритмі кожен об'єкт x_i розподіляється усіма кластерами з імовірністю $g_{iy} = P\{y_i = y\}$.

У алгоритмі k-середніх кожен об'єкт приписується лише одному кластеру, тому форма кластерів не формалізується. Загальна ідея алгоритмів полягає у мінімізації відстаней між об'єктами в кластерах. Зупинка відбувається, коли мінімізувати відстані більше вже неможливо. Функція, що мінімізується, в разі k-means має вигляд:

$$J = \sum_{k=1}^M \sum_{i=1}^N d^2(x_i, c_k) \quad (3)$$

де $x_i \in X$ - об'єкт кластеризації,

$c_j \in C$ - центр кластера, центроїд, $|X| = N, |C| = M$.

На момент старту алгоритму має бути відоме число C (кількість кластерів). В процесі аналізу поліграфічного виробництва кожен параметр повинен бути пов'язаним з p -мірним вектором. Даний вектор містить у собі оцінки параметрів. Кожен об'єкт описується вектором:

$$x_i = \{x_{i1}, \dots, x_{in}\}, x_i \in X^1 \quad (4)$$

де X^1 – сукупність векторів, що характеризує деякий кластер чи визначений напрямок (вектор).

Для проведення кластерного аналізу сукупності стратифікуючи векторів необхідне задання першочергових коректних умов кластеризації. Для виявлення цих умов потрібне першочергове сортування.

При проведенні ряду процедур кластеризації технологічного процесу поліграфічного виробництва формуються стійкі вектори – центри x_{in} . Основною метою цих індикаторів є пришвидшення процесу кластеризації за рахунок того, що у якості центрів кластерів будуть вибрані саме ці вектори. Отже, кожен параметр технологічного процесу може бути представлений у якості вектора, компонентами якого є характеристиками параметрів, що необхідні для його оцінки.

ВИСНОВОК

Розподілення технологічного процесу на кластери дозволяє розвантажити центральний процесор методом виключення з-під його розрахунків значної кількості інформації. Розподіленими системами легше управляти та вносити доповнення у вже існуючий процес. Крім того, за допомогою кластеру, ми отримуємо узагальнені параметри, що відображаються в загальній системі.

Управління локальними елементами відображається на основі диференціювання (оцінювання) параметрів взаємозв'язаних кластерів. У результаті використання методів кластеризації отримуємо підвищення якості поліграфічної продукції: обслуговуючий персонал РРМ лише вносить мінімальні корективи, оператор звільнений від аналізу ідентичних параметрів; основні ж операції контролюються автоматикою. Це дозволяє отримати більшу кількість поліграфічної продукції протягом одиниці часу.

СПИСОК ПОСИЛАНЬ

- [1] Веретенников В. І. Управління проектами / В. І. Веретенников, Л. М. Тарасенко, Г. І. Гевлик. — К.: ЦНЛ, 2006. — 280 с.
- [2] Самарин Ю. Н. Автоматизация управления полиграфическим предприятием / Ю. Н. Самарин, П. К. Иванов // КомпьюАрт. — 2006. — № 8. — С. 40–47.
- [3] Булычев Ю.Г. Бурлай И.В. Системный подход к моделированию сложных динамических систем в задачах оптимизации с прогнозирующей моделью // АИТ. – М.: Наука, 1996. – № 3. – С. 34-47.
- [4] Яковлев Ю.П. Контролинг на базі інформаційних технологій. – К.: ЦНЛ, 2006. – 318с.