

Partitionner un réseau de distribution d'eau potable en différentes zones de qualité d'eau

Nicolas Cheifetz¹, Pierre Mandel¹, Karine Delabre¹, Véronique Heim²

¹ Veolia Eau d'Ile-de-France, Le Vermont, 28 Boulevard de Pesaro, F-92751 Nanterre, France
{nicolas.cheifetz,pierre.mandel,karine.delabre}@veolia.com

² Syndicat des Eaux d'Ile-de-France, 120 Boulevard Saint-Germain, F-75006 Paris, France
v.heim@sedif.com

Mots-clés : *algorithme EM, partition par consensus, graphe d'un réseau hydraulique.*

1 Contexte industriel et problématique

Partitionner le réseau en différentes zones de qualité d'eau permet non seulement de mieux comprendre le fonctionnement du système de distribution en eau mais aussi d'aider tout détecteur d'événement à base de données [1, 5]. En effet, regrouper les points de mesure en réseau facilite l'identification des événements en qualité d'eau détectés. Deux approches ont été formulées récemment pour grouper les nœuds du graphe de distribution ayant une qualité d'eau similaire sur la base de traçages hydrauliques simulés [4, 6] i.e. une série temporelle de concentration est simulée pour chaque nœud et chaque traçage. Ces travaux appliquent un algorithme Kmeans classique sur les nœuds et souffrent de deux inconvénients : le non alignement des séries temporelles pour des nœuds distincts et le regroupement de nœuds avec des distances non similaires dans différents traçages. On propose dans cette contribution de répondre à ces deux défauts avec respectivement l'usage de la Dynamic Time Warping (DTW) entre des paires de séries de concentration et un partitionnement par consensus sur la base de plusieurs partitions indépendantes obtenues par un modèle probabiliste.

2 Partitionnement par consensus via une méthode approchée

La méthodologie proposée est décrite par trois étapes successives majeures, illustrée par la Figure (1), pour partitionner automatiquement les nœuds d'un système de distribution d'eau. La première étape consiste à créer via le logiciel Synergi WaterTM un traçage hydraulique à partir de chaque point d'entrée du réseau e.g. usine de production, station de pompage, réservoir, point d'échange avec opérateur voisin,... On obtient un jeu de données volumineux.

Pour chaque traçage, on définit $\mathbf{x}_i = (x_{i1}, \dots, x_{iT})$ comme la série temporelle de concentration du nœud i composée de T pas de temps. Soit n le nombre de nœuds dans le graphe réduit du modèle hydraulique avec la suppression des nœuds en forêt ainsi que les nœuds de degré 2, et on définit m le nombre de nœuds atteints pour le traçage étudié avec $m \leq n$. Inspirée par [2], la matrice de similarité suivante agrège les informations en un vecteur $(\mathbf{y}_1, \dots, \mathbf{y}_n)^T$:

$$\begin{pmatrix} 0 & * & \dots & * & * & \dots & * & * & * \\ * & \ddots & D(i,j) & \vdots & * & \vdots & * & SP(i) & \\ * & D(j,i) & \ddots & * & * & \vdots & * & \vdots & \\ * & * & * & 0 & * & \dots & * & SP(m) & \\ \hline * & \dots & \dots & * & 0 & \dots & 0 & \inf & \\ * & \vdots & \vdots & * & 0 & \vdots & 0 & \vdots & \\ * & \dots & \dots & * & 0 & \dots & 0 & \inf & \end{pmatrix}_{n \times n+1} = \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_n \end{pmatrix} \quad (1)$$

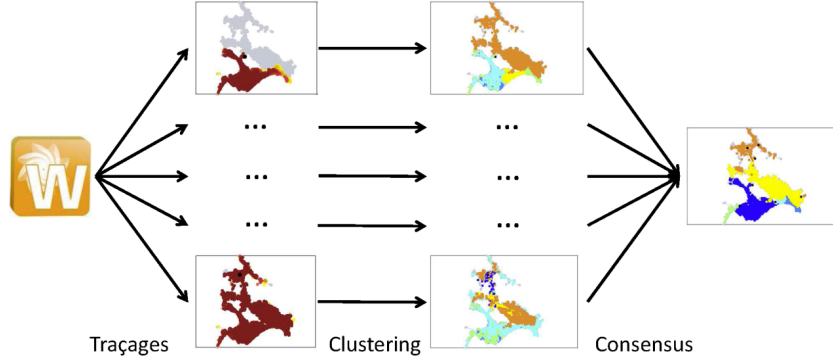


FIG. 1 – Méthodologie proposée pour partitionner le réseau de distribution.

Avec $D(i, j)$ une mesure $DTW(\mathbf{x}_i, \mathbf{x}_j)$ normalisée, $SP(i)$ un plus court chemin normalisé depuis le point d'entrée du traçage jusqu'au nœud i , et $1 \leq i, j \leq m$. La densité de probabilité d'un nœud i est exprimée par un modèle de mélange gaussien $p(\mathbf{y}_i; \boldsymbol{\theta})$ et on note $\tau_{ik}(\hat{\boldsymbol{\theta}})$ la probabilité a posteriori d'appartenance du nœud i au cluster k . La seconde étape consiste à lancer un algorithme de type Expectation Maximization (EM) pour estimer le vecteur paramètre $\hat{\boldsymbol{\theta}}$ de chaque traçage. Il est à noter que le nombre de clusters est fixé en minimisant un critère de type BIC, et on initialise chaque algorithme EM avec une partition spatiale sur les nœuds (i.e. Kmeans appliqué aux plus courts chemins) ce qui conduit à de bonnes solutions en pratique.

Enfin, on définit une mesure (pseudo-métrique) de dissimilarité entre deux nœuds i et j : $1 - \frac{1}{S} \sum_{s=1}^S \sum_{k=1}^{K_s} \tau_{ik}(\hat{\boldsymbol{\theta}}) \cdot \tau_{jk}(\hat{\boldsymbol{\theta}})$, avec τ une probabilité a posteriori et S le nombre de traçages simulés. La troisième étape consiste à regrouper les nœuds similaires à partir de cette mesure par un algorithme de clustering hiérarchique avec lien complet [3]. Une stratégie de propagation de labels est alors utilisée pour étiqueter les nœuds hors du graphe réduit, et on obtient une unique partition sur tous les nœuds du réseau interprétable en matière d'origine de l'eau.

3 Conclusions

Ce papier propose une méthode de partitionnement d'un réseau de distribution d'eau potable. Les résultats expérimentaux obtenus sur un graphe d'environ 20 000 nœuds et 22 000 arrêtes montrent la pertinence et une certaine flexibilité de l'approche formulée.

Références

- [1] Eustace M Dogo, Nnamdi I Nwulu, Bhokisipho Twala, and Clinton Aigbavboa. A survey of machine learning methods applied to anomaly detection on drinking-water quality data. *Urban Water Journal*, 16, 2019.
- [2] Issam Falih, Nistor Grozavu, Rushed Kanawati, and Younès Bennani. ANCA : Attributed Network Clustering Algorithm. In *International Conference on Complex Networks and their Applications*, volume 689, pages 241–252. Springer, 2017.
- [3] Arno Fritsch, Katja Ickstadt, et al. Improved criteria for clustering based on the posterior similarity matrix. *Bayesian analysis*, 4(2) :367–391, 2009.
- [4] Pierre Mandel, Marie Maurel, and Damien Chenu. Better understanding of water quality evolution in water distribution networks using data clustering. *Water research*, 87, 2015.
- [5] Fitore Muharemi, Doina Logofătu, and Florin Leon. Machine learning approaches for anomaly detection of water quality on a real-world data set. *Journal of Information and Telecom*, 2019.
- [6] Tian Qin and Dominic L Boccelli. Grouping water-demand nodes by similarity among flow paths in water-distribution systems. *Journal of Water Resources Planning and Management*, 2017.