Chapter 1

Présentation du projet

1.0.1 Presentation de ENGIE

1.0.2 But

A partir des données de signaux électriques en provenance des panneaux solaires, il s'agit de détecter les sous-performances des équipements et d'analyser les empreintes des différentes anomalies. Ces informations seront transmises à l'exploitant. Dans un second temps, selon l'avancement, il faudrait aussi identifier les causes des perturbations. Surveiller nos équipements et détecter les anomalies qui illustrent la sous performance des équipements.

1.1 Introduction scientifique

1.2 Etude bibliographique sur la détection d'anomalies pour un système transient

En analysant la littérature concernant

1.2.1 LSTM

1.2.2 Décomposition en Wavelet + réseau de neuronne

1.2.3 Random forest

1.3 Methodologie

1. Analyser les données et décomposer les résidus pour retrouver les différents effets

    - Approche par ondelettes : décomposition du signal sur la base de signatures typiques du signal

2. Approche globale pour retrouver les effets de façon simultanée

Chapter 2

Description des données

2.1 Influence du rayonnement solaire sur la production électrique

2.2 Différents modèles de production électriques

2.3

Chapter 3

Détection d'anomalie

3.1 Présentation générale des différentes méthodes

3.1.1 Types d'anomalies

La détection d'anomalie dans les données est une méthode consistant à trouver des motifs non usuels dans la donnée. Ces anomalies sont différentes du bruit existant dans les données. Il existe différent types d'anomalies:

1. Les anomalies ponctuelles (un point dont la valeur excède la gamme de valeur de la distribution des données)

2. Les anomalies contextuelles (des valeurs non usuels pour le contexte donné)

3. Les anomalies collectives (une gamme de valeur non usuels)

Dans le cas où on suppose que les données sont paramétriques, on suppose que les données normales suivent une loi gaussienne. Les anomalies sont les données qui ne suivent pas cette distribution.

L'anormalité est caractérisée par la distance de Mahalanobis

$$d(x, y) = \sqrt{(x-y)^\top \Sigma (x-y)} \tag{3.1}$$

Si on ne suppose rien sur la loi que suit les données normales, d'autres méthodes sont utilisées comme

1. Density based

2. Distance based

Random forest

On définira une notion de similarité et de distance entre les données .

La similarité entre les 2 observation est basé sur le nombre de fois les 2 observations correspondent à la même feuille.

$$prox(n, k) = \frac{similarity}{numbertree} \tag{3.2}$$

Les anomalies correspondent aux observations qui ont une petite pro par rapport aux autres observations

3.1.2 Isolation Forest

Un arbre est construit pour détecter l'anomalie

3.1.3 LOF (Local Outlier Factor)

Identifier la densité local de chaque point et la comparer avec les voisins. The local reachibility factor density is computed as

$$density = .. \tag{3.3}$$

3.1.4 1 class SVM

3.1.5 L'algorithme du LSTM

3.1.6 Extraction des coefficients des ondelettes à utiliser comme feature

Les données peuvent être projetés dans une autre base pour pouvoir détecter le comportement normal et anormal. Dans le cas des séries temporelles, on utilise souvent une base composée des coefficients d'ondelettes.

Bibliography

[1] BREIMAN, Leo. *Random forests.* Machine learning, 2001, vol. 45, no 1, p. 5-32.

[2] BREUNIG, Markus M., KRIEGEL, Hans-Peter, NG, Raymond T., et al. *LOF: identifying density-based local outliers.* In : ACM sigmod record. ACM, 2000. p. 93-104.

[3] CHANDOLA, Varun, BANERJEE, Arindam, et KUMAR, Vipin. *Anomaly detection: A survey.* ACM computing surveys (CSUR), 2009, vol. 41, no 3, p. 15.

[4] MALHOTRA, Pankaj, VIG, Lovekesh, SHROFF, Gautam, et al. *Long short term memory networks for anomaly detection in time series.* In : Proceedings. Presses universitaires de Louvain, 2015. p. 89.

[5] KEOGH, Eamonn, LONARDI, Stefano, et CHIU, Bill'Yuan-chi'. *Finding surprising patterns in a time series database in linear time and space.* In : Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2002. p. 550-556.

[6] Li, Z., Li, Z., Yu, N., Wen, S. *Locality-Based Visual Outlier Detection Algorithm for Time Series* Security and Communication Networks, 2017.

[7] Chen, C., Liu, L. M. (1993). *Joint estimation of model parameters and outlier effects in time series.* Journal of the American Statistical Association, 88(421), 284-297.

[8] DING, Hui, TRAJCEVSKI, Goce, SCHEUERMANN, Peter, et al. *Querying and mining of time series data: experimental comparison of representations and dis-*

tance measures. Proceedings of the VLDB Endowment, 2008, vol. 1, no 2, p. 1542-1552.

Chapter 4

Decomposition en ondelettes

4.1 Principe

Pour le traitement du signal, différentes méthodes de décomposition existent. La plus connue et utilisée, est la transformée de Fourier. Cette dernière permet de filtrer le signal à différentes fréquences temporelles. Elle informe donc sur le contenu fréquentiel du signal.

Elle possède une limitation majeure dans la connaissance du comportement local d'une fonction. Ce défaut a été observé par D.Gabor qui a réussi à pallier à cette inconvénient en utilisant la STFT (short-time Fourier transform). Seulement la résolution temporelle restait insuffisante. Elle n'est donc pas adéquate pour l'étude de signaux dans un régime transitoire. [9]

La décomposition en ondelettes permet de pallier à ce défaut. Elle prend en compte la variabilité fréquentielle de la fonction. Elle donne donc une *représentation temps-fréquence d'un signal* [2]. et permet donc de fournir une localisation temporelle.

4.1.1 Definition

La transformée en ondelettes [2] continue d'une fonction $f \in L^2(\mathbb{R})$ au point $x \in \mathbb{R}$ et à l'échelle $s > 0$ est définie par :

$$W_f(x,s) =< f, \Psi_{x,s} >= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{s}} f(t) \Psi(\frac{t-x}{s}) dt \qquad (4.1)$$

Une transformée en ondelettes peut s'écrire sous la forme d'un filtrage par convolution :

$$W_f(x,s) = f * \psi_s^*(x) \qquad (4.2)$$

avec $\psi_s^*(x) = \frac{1}{\sqrt{s}} \psi(\frac{-t}{s})$

La transformée en ondelette satisfait les propriétés de conservation de l'énergie du signal.

L'ondelette choisie doit avoir :

1. un support compact pour obtenir une localisation dans l'espace

2. une moyenne égale à 0

propriété: orthogonalité, lissante

Il existe plusieurs types d'ondelettes :

1. Haar

2. Mexican hat

3. Morlet

4. Daubechie

On utilisera l'ondelette Daubechie qui est la plus souvent utilisée d'après la littérature.

La transformée en ondelette permet de représenter une fonction comme une combinaison linéaire de fonctions de bases d'ondelettes.

Elle permet de réprésenter une fonction dans une *représentation multirésolution d'ondelettes*. Celà consiste à une séquences d'espaces fermées V_m de $L^2(\mathbb{R})$ où $L^2(\mathbb{R})$ est un espace d'Hilbert de outes les fonctions au carré intégrables. Ces espaces caractérisent le comportement d'une fonction à l'échelle 2^m échantillons par unité de longueurs.

Les coefficients de la transformée en ondelettes sont calculées par

$$W_f(x, s) = f * \psi_s^*(x) \tag{4.3}$$

4.2 Méthodologie

4.2.1 Decomposition

La fonction que l'on souhaite décomposer est réalisée de la façon suivante :

$$f(x) = \sum_{k=0}^{2j0-1} \tag{4.4}$$

$$f(x) = f_{J0}(x) + \sum_{j>j0} D_j(x) \tag{4.5}$$

Les coefficients d'ondelettes sont calculées de la manière suivante : avec le coefficient d'échelle :

$$\alpha = \int_0^1 f(x)\phi(x)dx \tag{4.6}$$

et le coefficient de détail :

$$\beta_{j,k} = \int_0^1 f(x)\psi(x)dx \tag{4.7}$$

$phi(x)$ est appelé l'ondelette mère et $\psi(x)$ est l'ondelette père .

Pour une ondelette de Haar (ou Daubechie) , elles sont de la forme

$$\psi(x) = \begin{cases} -1 & \text{si } \{x\} \in [0, 1/2] \\ 1 & \text{si } \{x\} \in]1/2, 1] \end{cases} \tag{4.8}$$

$$\phi(x) = \begin{cases} 1 & \text{si } \{x\} \in [0, 1[\\ 0 & \text{sinon.} \end{cases} \tag{4.9}$$

4.2.2 Détection d'anomalie à l'aide des coefficients dans la base en ondelettes

Pour détecter les anomalies, on se concentrera sur la projection des données sur la base formée des coefficients d'ondelettes.

4.3 Résultat

Bibliography

[1] Stéphane Mallat, *A Wavelet Tour of Signal Processing.* Elsevier, 1999.

[2] Jérémie Bigot, *Analyse par ondelettes* Université Paul Sabatier, 2009

[3] Amara Graps *An Introduction to Wavelets* IEEE computational science and engineering, 1995, vol. 2, no 2, p. 50-61, 1995.

[4] Charles K Chui *An Introduction to Wavelets* Elsevier, 2016.

[5] GENÇAY, Ramazan, SELÇUK, Faruk, et WHITCHER, Brandon J *An introduction to wavelets and other filtering methods in finance and economics* Elsevier, 2001.

[6] WOJTASZCZYK, Przemyslaw. *A mathematical introduction to wavelets.* Cambridge University Press, 1997.

[7] STRANG, Gilbert et NGUYEN, Truong. *Wavelets and filter banks.* SIAM, 1996.

[8] MALLAT, Stephane et HWANG, Wen Liang. *Singularity detection and processing with wavelets. IEEE transactions on information theory* X vol. 38, no 2, p. 617-643, 1992.

[9] Gao, R. X., Yan, R. (2010) *Wavelets: Theory and applications for manufacturing.* Springer Science & Business Media.

[10] GULER, Inan et UBEYLI, Elif Derya *ECG beat classifier designed by combined neural network model.* Pattern recognition, 2005, vol. 38, no 2, p. 199-208.

[11] INCE, Turker, KIRANYAZ, Serkan, et GABBOUJ, Moncef. *A generic and robust system for automated patient-specific classification of ECG signals.* IEEE Transactions on Biomedical Engineering, 2009, vol. 56, no 5, p. 1415-1426.