

L'intelligence artificielle au service de la configuration des actifs

26 février 2021, 01:00

La connexion d'une centrale photovoltaïque à une plate-forme de suivi des ressources d'énergie renouvelable requiert une configuration précise. Cette activité manuelle, chronophage et sujette aux erreurs peut être appuyée par une approche reposant sur les données.

L'entreprise 3E basée à Bruxelles se fixe pour objectif une durabilité maximale. Ses experts et ingénieurs aident les acteurs industriels à optimiser la production et l'utilisation des énergies (renouvelables) en proposant des services innovants de développement logiciel et de consultance. 3E peut se targuer d'une longue tradition de l'innovation dans les domaines des installations d'énergie renouvelable, des systèmes d'alimentation intelligente et de l'amélioration des processus en matière d'énergie dans le monde entier.

3E propose la plate-forme SynaptiQ de suivi des ressources d'énergie renouvelable, telles que les centrales éoliennes et photovoltaïques (PV). La connexion d'une centrale PV à SynaptiQ pose un défi particulier : l'intégration. Il s'agit de configurer le processus d'intégration des données pour cette centrale ainsi que son jumeau numérique pour pouvoir calculer correctement les indicateurs clés de performance. Selon la taille, la complexité et les caractéristiques propres d'une centrale, l'intégration peut être un long processus nécessitant l'intervention d'ingénieurs expérimentés. Voici les principales sources de complexité pour les ingénieurs :

- une normalisation insuffisante avec un large éventail de fabricants de matériel, de protocoles de communication, de formats de données et de configurations ad-hoc de systèmes SCADA sur site
- une documentation souvent incomplète (ou même manquante !) sur la topologie, les types d'appareils et les interconnexions d'une centrale.

Par conséquent, la configuration de certaines des plus grandes centrales nécessite souvent une étude détaillée et une analyse rigoureuse des données recueillies, réparties sur des centaines d'appareils, pour reconstruire la configuration. Vu la forte demande de nouvelles centrales à connecter, on ne configure parfois que les indicateurs de performance les plus importants, tandis qu'on ignore les indicateurs de moindre pertinence, surtout si leur coût de configuration est élevé.

Le moteur SynaptiQ ETL met déjà en œuvre un système de suivi des erreurs très complet, qui enregistre en permanence les indicateurs inconnus ou incorrects reçus d'une centrale (p. ex. une discordance entre les données reçues et la configuration jumelle numérique), ce qui est d'une grande utilité pour l'équipe d'ingénieurs.

Un projet de recherche a été mis sur pied avec l'EluciDATA Lab de Sirris afin d'exploiter davantage ces informations sur les erreurs. L'objectif est de réduire le temps de configuration et donc

d'augmenter la couverture des indicateurs, grâce à une approche reposant sur les données qui recommande pour une centrale PV donnée les paramètres de configuration à fournir et leurs valeurs.

Nécessité d'un prétraitement rigoureux des données

Les données fournies par 3E couvrent une population très diverse de plusieurs milliers de centrales PV, certaines étant très similaires et d'autres très différentes. La variabilité de chaque configuration de centrale va de quelques paramètres à des dizaines de paramètres : pour certaines centrales, des paramètres et valeurs standard suffisent, tandis que pour d'autres, plus de 10 paramètres sont spécifiés.

Il est important de noter que l'occurrence des configurations uniques est très déséquilibrée. Alors que certaines configurations ne se présentent qu'une fois, d'autres dominent l'ensemble des données en revenant plus de 2.500 fois. Il est donc important d'équilibrer d'abord le rapport des configurations uniques de sorte à accorder plus de poids aux configurations fréquentes, mais sans ignorer les moins fréquentes. Nous calculons donc le poids m d'une configuration comme suit :

$$m = \begin{cases} 2, & \text{if } \log_2(\#n) < 2 \\ 100, & \text{if } \log_2(\#n) > 100 \\ \log_2(\#n), & \text{else} \end{cases}$$

où n est le nombre d'occurrences d'une configuration.

Extraction de règles d'association

Pour recommander une configuration intéressante, nous utiliserons ce qu'on appelle les « association rules ». Dans la première étape vers la prédiction d'une configuration, des soi-disant « frequent itemsets » ont été calculés à partir de l'ensemble équilibré de configurations, qui regroupent les différents paramètres et valeurs dans des listes d'éléments qui se présentent (plus) souvent ensemble. Le rapport entre le nombre de fois où les paramétrages uniques se sont présentés ensemble et l'ensemble de données est représenté par le « support ». Plus le support est important, plus la cooccurrence de plusieurs paramètres et de leurs valeurs correspondantes est évidente. Cela donne une liste de plus de 18.000 éléments fréquents et leur support.

support		itemsets
1367	0.240602	(csv.timestamp.column__1, csv.separator__, serialid.pattern.group__1)
4611	0.159398	(csv.header.line__1, csv.timestamp.column__1, csv.separator__)
180	0.154887	(timestamp.pattern__yyyy-MM-dd HH:mm:ss, csv.timestamp.column__1, csv.separator__)
178	0.132331	(timestamp.pattern__yyyy-MM-dd HH:mm:ss, csv.timestamp.column__1, serialid.pattern.group__1)
4612	0.120301	(csv.header.line__1, csv.timestamp.column__1, serialid.pattern.group__1)
...	...	...

Dans un deuxième temps, nous appliquons l'algorithme d'extraction des règles d'association à l'ensemble des paramétrages fréquents obtenus précédemment, afin d'identifier des règles basées

sur les configurations courantes. Cela donne un ensemble de règles indiquant les paramètres et leurs valeurs correspondantes qui vont ensemble. Pour chaque règle, on peut effectuer plusieurs mesures d'*ajustement*, c'est-à-dire des mesures de la force d'un rapport. En ne conservant que les règles pertinentes, on a quand même obtenu plus de 700.000 règles.

De ces règles, on peut dériver des prédictions de configuration pour chaque centrale :

Input settings	recommendations	goodness measures				
antecedents	consequents	antecedent support	consequent support	support	confidence	lift
(csv.header.line__1, csv.timestamp.column__1, dynamic.units__false, tag.pattern.group__3)	(csv.data.line__2, serialid.pattern.group__2)	0.006015	0.006015	0.006015	1.0	166.25
(csv.header.line__1, csv.timestamp.column__1, csv.data.line__2, serialid.pattern.group__2)	(dynamic.units__false, tag.pattern.group__3)	0.006015	0.006015	0.006015	1.0	166.25
(csv.header.line__1, dynamic.units__false, serialid.pattern.group__2, csv.data.line__2)	(tag.pattern.group__3, csv.separator__3)	0.006015	0.006015	0.006015	1.0	166.25
(csv.header.line__1, dynamic.units__false, tag.pattern.group__3, csv.separator__3)	(csv.data.line__2, serialid.pattern.group__2)	0.006015	0.006015	0.006015	1.0	166.25
(csv.header.line__1, csv.data.line__2, serialid.pattern.group__2, csv.separator__3)	(dynamic.units__false, tag.pattern.group__3)	0.006015	0.006015	0.006015	1.0	166.25

Évaluation des résultats

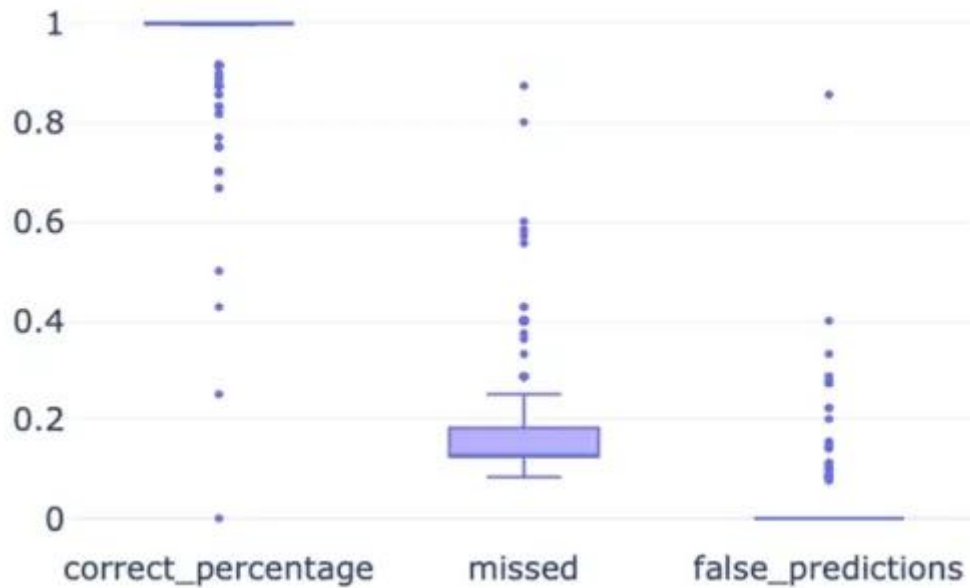
L'objectif était de fournir des recommandations pour les centrales pour lesquelles des erreurs d'importation des données avaient été constatées. Trois raisons peuvent expliquer ces erreurs :

- Une ou plusieurs valeurs d'un paramètre sont mal définies.
- Un ou plusieurs paramétrages manquent.
- Un ou plusieurs paramétrages sont inutiles et trompeurs.

C'est pourquoi, pour chaque centrale, plusieurs sous-ensembles de paramètres et leurs valeurs correspondantes ont été utilisés comme variables d'entrée dans le moteur de recommandation. Les sous-ensembles comprennent toutes les combinaisons de 2 à $k - 1$ paires paramètre-valeur, k étant le nombre de paramètres défini dans la configuration d'origine. Soulignons que le nombre de sous-ensembles croît de manière exponentielle avec k et qu'en fonction des paramètres d'entrée, plusieurs paramètres différents et pour chaque valeur différente pourraient être recommandés par sous-ensemble. Afin d'arrêter les paramètres et valeurs de paramètre définitifs, un vote à la majorité a été appliqué : l'ensemble de paramètres et valeurs correspondantes le plus recommandé dans tous les sous-ensembles a été retenu. En cas d'égalité, on a choisi celui qui avait le meilleur taux de confiance.

Dans l'ensemble, de bons résultats ont été obtenus :

- Le pourcentage moyen de valeurs de paramètre correctement prédites est d'environ 90%.
- Moins de 20% de paramètres manqués en moyenne, c'est-à-dire paramètres qui n'étaient pas prévus dans la configuration.
- Moins de 10% de fausses prédictions des valeurs des paramètres en moyenne.



Conclusion

Les résultats montrent qu'une approche reposant sur l'intelligence artificielle peut effectivement être utilisée pour appuyer la configuration des ressources. Nous pouvons fournir des recommandations sur les paramètres de configuration qui doivent être fournis et leurs valeurs.

Voici toutefois les points importants à retenir :

- Un prétraitement rigoureux des données est essentiel pour obtenir de bons résultats : introduire simplement l'ensemble des données tel quel dans l'algorithme donnerait des recommandations (erronées) dominées par des configurations plus fréquentes.
- Une procédure spécifique doit être définie pour sélectionner la recommandation avec le meilleur taux de confiance : l'algorithme d'extraction de règles génère un grand nombre de règles, pouvant mener à des recommandations contradictoires.