

Présentation du sujet

La prédiction de la production de futurs puits est une problématique importante. Être capable d'anticiper la performance d'un puits va permettre de planifier les investissements à long terme et d'optimiser les stratégies d'acquisition et de vente de terrain. Usuellement, l'industrie pétrolière s'appuie sur une procédure standard, la Decline Curve Analysis (DCA

http://petrowiki.org/Production_forecasting_decline_curve_analysis).

Cette approche générale ne tient pas compte des spécificités de chaque puits. Un groupe pétrolier cherche à aller plus loin, en s'appuyant sur des modèles plus précis et proches de la réalité. Pour cela, un programme de récolte de données permettant de caractériser la physique de chaque puits a été mis en place. Sur plusieurs années, des informations de géolocalisation, production, géologie ont été extraites et nettoyées. L'objectif est désormais d'utiliser au mieux ces données. Plus particulièrement, vous allez être amené à prédire deux indicateurs de performance : les quantités de gaz (GAS360) et de condensat (CUM360) extraits d'un puits, un an après sa mise en service, c'est-à-dire une fois que le système de production est stabilisé. Pour rajouter une dimension probabiliste au problème, il vous est demandé de ne prédire non pas deux valeurs uniques par puits, mais une surface de confiance (bornes hautes et basses pour le gaz ainsi que le condensat). Cette surface devra être la plus petite possible, tout en contenant les vraies valeurs observées.

Description des données

Deux fichiers seront fournis.

1. Un fichier de données d'**apprentissage** avec 360 observations. La première colonne contient les identifiants des puits et ne servira pas pour la prédiction. Les colonnes D, E, F contiennent des dates de l'historique du puits. Les autres colonnes contiennent des données physiques **normalisées**. Les deux dernières colonnes contiennent les productions observées de gaz et d'oil GasCum360, OilCum360.
2. Un fichier de données de test ayant exactement la même structure que le fichier d'apprentissage hormis les deux dernières colonnes. Ce deuxième fichier servira à évaluer votre travail avec la métrique décrite ci-dessous.

Valeurs manquantes. Les fichiers de données contiennent des valeurs manquantes dont il faut trouver une méthode pour les compléter.

Métrique d'évaluation

Le résultat de votre travail sera un tableau ayant le format suivant :

API	Gas360_INF	Gas360_SUP	Oil360_INF	Oil360_SUP
661	0.42960043	0.72960043	-1.57039957	-1.27039957
309	-0.11311024	0.18688976	-2.11311024	-1.81311024
382	-2.31561571	-2.01561571	-4.31561571	-4.01561571
...

Donc, pour chaque puits identifié par son API, il faut fournir les bornes min et max de l'estimation du GAS et du OIL. Pour chaque puit p , on calculera la surface

$$S(p) = (Gas360_SUP(p) - Gas360_INF(p)) * (Oil360_SUP(p) - Oil360_INF(p))$$

Et une variable binaire $in(p) = 1$ si la prédiction

$$GasCum360(p) \in [Gas360_INF(p), Gas360_SUP(p)] \text{ et } OilCum360(p) \in [Oil360_INF(p), Oil360_SUP(p)]$$

et $in(p) = 0$ sinon. Il s'agira alors de **minimiser** la métrique suivante :

$$M = \frac{1}{nbpuits} \sum_{p=1}^{nbpuits} (S(p)in(p) + 10(1 - in(p)))$$

Codage

Les modèles enseignés en cours seront utilisés pour prédire les intervalles mentionnés dans la section Métrique d'évaluation. Le langage de programmation sera Matlab (Octave) ou Python (qui ouvre le droit à un bonus). Le meilleur résultat correspondra à la métrique minimum.

Rapport. Un rapport de 5 pages décrira les 4 points suivants :

1. Présentation de la méthode utilisée
2. Présentation du raisonnement qui vous a amené à développer cette méthode
3. Description de l'implémentation logicielle, avec les éléments de code importants
4. Améliorations possibles

La note finale sera le résultat des deux notes : note sur la métrique ; note sur le rapport.

Format de l'envoi

M'envoyer par mél le travail fourni dans un fichier compressé de nom :

Projet_Nom1_Nom1.zip

Date limite. 26 février 2018.