

Projet Cow-neck-TID

(Traitement Intelligent des Données du cou des vaches)
avec ITK

Projet en équipe de 2 à 3 personnes.

Pour des raisons logistiques, ce projet est actuellement à date fixe jusqu'au 1er décembre 2021.

1. Avant-propos

Dans le cadre du programme Agritech entre Hectar et 42, le projet Cow-neck-TID est effectué en collaboration avec ITK (<https://itk.fr>).

2. Contexte agricole

Faute de ne pouvoir raisonnablement passer 100% de son temps avec son bétail, il est envisagé d'effectuer un suivi partiellement automatisé de son cheptel. Le campus Hectar dispose de vaches connectées : chaque animal porte autour de son cou un collier connecté qui permet de le localiser, mais également de suivre les différents mouvements de sa tête.

3. Partie 1 : traitement et visualisation de la donnée

Vous disposez d'informations en provenance des capteurs sur tout un historique qui permettra de faire apprendre vos algorithmes de machine learning (en partie 2), et également en temps réel (ou presque) pour les vaches du campus Hectar. Vous devez :

- filtrer la donnée afin d'enlever les éléments aberrants, principalement sur les données temps réel. Les données historiques d'entraînement sont quant à elles déjà propres.
- effectuer des représentations visuelles de l'ensemble de ces données (data visualisation avec des histogrammes, camemberts, courbes, cartes, etc...) de sorte à permettre une lecture simple à un exploitant pour ses animaux. Selon que vous travaillez sur de la donnée historique qui contient déjà les analyses et calculs de l'état de l'animal, ou bien temps réel, toutes les informations, courbes, etc... ne pourront pas forcément être affichées.
- mettre en évidence les cycles, les éléments récurrents, les changements, des événements inhabituels, des ruptures d'habitudes.

Le rendu se fait sous la forme d'une plateforme web compatible desktop et mobile. Un conteneur apparaît comme un choix naturel pour ce rendu.

4. Partie 2 : prédiction comparée

ITK dispose d'un système de prédiction de l'activité de chaque vache sur la base de ces données. Ce système permet notamment de connaître l'état de la vache (rumine, dort, ...) mais également de savoir si un vêlage est imminent ou si l'animal présente un comportement de chaleurs. Le chapitre 6 décrit les situations analysées par le système de prédiction, dont vous retrouvez les résultats sur le jeu de données historique. Il vous faut réaliser un programme qui permet de retrouver ces mêmes prédictions, entraîné sur l'historique, puis appliqué sur les données quotidiennes des vaches du campus Hectar. Une vérification directe de vos prédictions sur le terrain semble appropriée pour confirmer la pertinence de vos algorithmes.

Le rendu se fait sous la forme d'un script en ligne de commande qui se base sur l'identifiant unique (ou les identifiants unique) de la vache et qui permet d'afficher l'activité probable liée au jeu de données testé. Ce mode de rendu permet une possible compétition entre groupes de projet. Un affichage du résultat de ce script doit également être effectué sur la page de visualisation des données créée en partie 1 (pour une vache donnée ou bien pour l'ensemble du cheptel). Sur le site web rendu, il est possible d'activer / désactiver l'utilisation de ce script, et donc faire apparaître / disparaître les courbes de résultats trouvés par le script.

5. Partie 3 : prédiction avancée

Dans cette partie il vous faut explorer de nouvelles pistes pour détecter d'autres situations ou comportements : rendement de la production de lait, maladie, stress, croissance de l'animal, cycles d'alimentation et de sommeil, ... Il est autorisé d'introduire de la data publique complémentaire.

A nouveau, il s'agit d'un script utilisable en ligne de commande qui doit être fourni, et de son intégration web cote-à-cote avec celui de la partie 2.

6. Considérations techniques et légales

Les différents protocoles de données sont disponibles à l'adresse suivante :

[tps://cloud.itkweb.fr/index.php/s/TKiQ2jQY6wAa9qs](https://cloud.itkweb.fr/index.php/s/TKiQ2jQY6wAa9qs)

Mot de passe : demandez le mot de passe à votre responsable 42 sur site, après signature de l'engagement de confidentialité.

ATTENTION : reportez-vous à la section "6.2 Considérations légales" avant d'accéder, télécharger, et utiliser ces données.

Un protocole de données est un jeu de données de référence permettant de qualifier les performances d'un algorithme de prédiction ou de détection. Dans le cas échéant, vous avez

à votre disposition quatre protocoles de données qui nous ont permis de qualifier les algorithmes de détection de vêlage et de chaleur.

Chaque protocole de données possède:

1. Des évènements de référence, c-a-d un comportement de l'animal observé directement sur le terrain ou détecté de manière fiable.
2. Pour chaque évènement, une série de mesures brutes provenant des capteurs Axel.
3. Ainsi qu'une série de données normalisées représentant l'activité de l'animal.

Les protocoles de données fournissent les évènements de références suivants:

Protocole libvel_laitier_ms_utc et libvel_laitier_prod_tvax

Définissent des évènements de type **tv_expulsion**. Cet événement indique une rupture de la poche des eaux de l'animal, signalant que le vêlage est imminent. Nous considérons généralement que le vêlage peut intervenir entre 1h et 2h après.

Protocole heatlive_allaitant_ms

Définis des évènements de type **heat_visual**. Cet événement indique qu'un éleveur a observé un comportement de chaleur.

Protocole heatlive_allaitant_prod_ia

Définit des évènements de type **artificial_insemination**. Cet événement indique qu'une insémination artificielle a été pratiquée sur l'animal. Cela indique qu'il devrait y avoir dans les données des signes de comportement de chaleur 12 heures avant cet événement.

6.1 Documentation des données et format.

La structure des fichiers de chaque protocole est organisé comme suit :

```
<nom du protocol>/  
|-- metadata.csv  
|-- ax_raw /  
---- FR1448330091_2018-11-27.parquet  
---- ...
```

```
| -- normalized/  
| ---- FR1448330091_2018-11-27.parquet
```

Metadata.csv

Ce fichier contient tous les événements de référence du protocole de données ainsi que des informations supplémentaires sur l'animal associé à cet événement. Les colonnes principales de ce fichier sont les suivantes:

event_id: Identification unique de l'événement. Cet identifiant sert principalement à faire l'association entre l'événement et les séries de mesures brutes et normalisées.

ref_event_datetime: Date de l'événement de référence

farm_id: identification de l'exploitation

cow_id: identification de l'animal.

country: Pays de l'exploitation

birth_day: Date de naissance de l'animal

breed: Race de l'animal

rank: type d'élevage, *laitier* ou *allaitant*

cattle_mode: *genisse* si l'animal a déjà vêlé, sinon *vache*

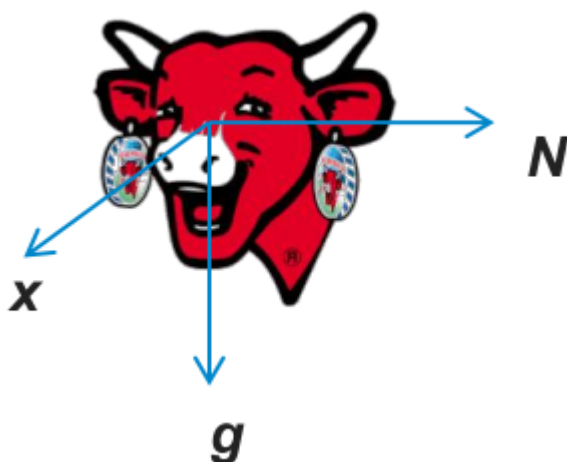
season: *summer*, *automne*, *winter*, *spring*

Dossier ax_raw

Ce dossier contient toutes les séries de données brutes provenant des capteurs Axel.

Chaque fichier, nommé `ax_raw/<event_id>.parquet` correspond à une série temporelle associée à un événement présent dans les métas-données.

Ces fichiers sont sous le format parquet pouvant être lu directement avec la librairie pandas: `pd.read_parquet.`



Les capteurs mesurent l'accélération dans un système de coordonnées (**x, g, N**) centré sur la tête de l'animal. La fréquence initiale d'échantillonnage est de 0.1 Hz (période 10 secondes), sous-échantillonné à **une période de 5 minutes**.

Dans ces fichiers de mesures, nous avons 9 variables toute les 5 minutes:

EAn: moyenne de l'accélération sur l'axe **N** sur les 5 minutes.

EAg: moyenne de l'accélération sur l'axe **g** sur les 5 minutes.

Egx: moyenne de l'accélération sur l'axe **x** sur les 5 minutes.

ChAn: somme des valeurs absolues de l'accélération sur axe **N** sur les 5 min

ChAg : somme des valeurs absolues de l'accélération sur axe **g** sur les 5 min

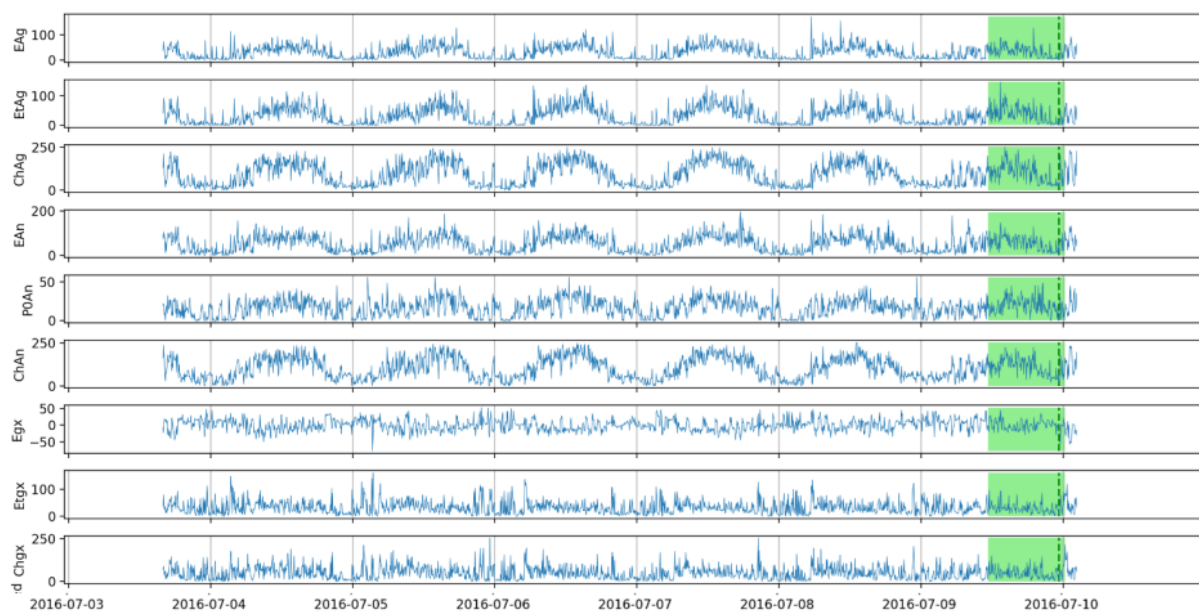
Chgx : somme des valeurs absolues de l'accélération sur axe **x** sur les 5 min

EtAg : écart-type de l'accélération sur axe **g** sur les 5 min

Etgx: écart-type de l'accélération sur axe **x** sur les 5 min

P0An : nombre de passages par 0 de l'accélération sur axe **N** sur les 5 min

Exemple de mesures : (9 variables, période de 5 min, 7 jours avant chaleurs)



Dossier normalized

Ce dossier contient toutes les séries dites "normalisées", représentant l'activité de l'animal suivant deux groupes de classe.

1. Classification du **comportement** de l'animal en 5 classes:

Rumination: *True* si l'animal rumine

Ingestion: *True* si l'animal est en ingestion

Rest: *True* si l'animal est en repos

Over Activity: *True* si l'animal est en sur-activité

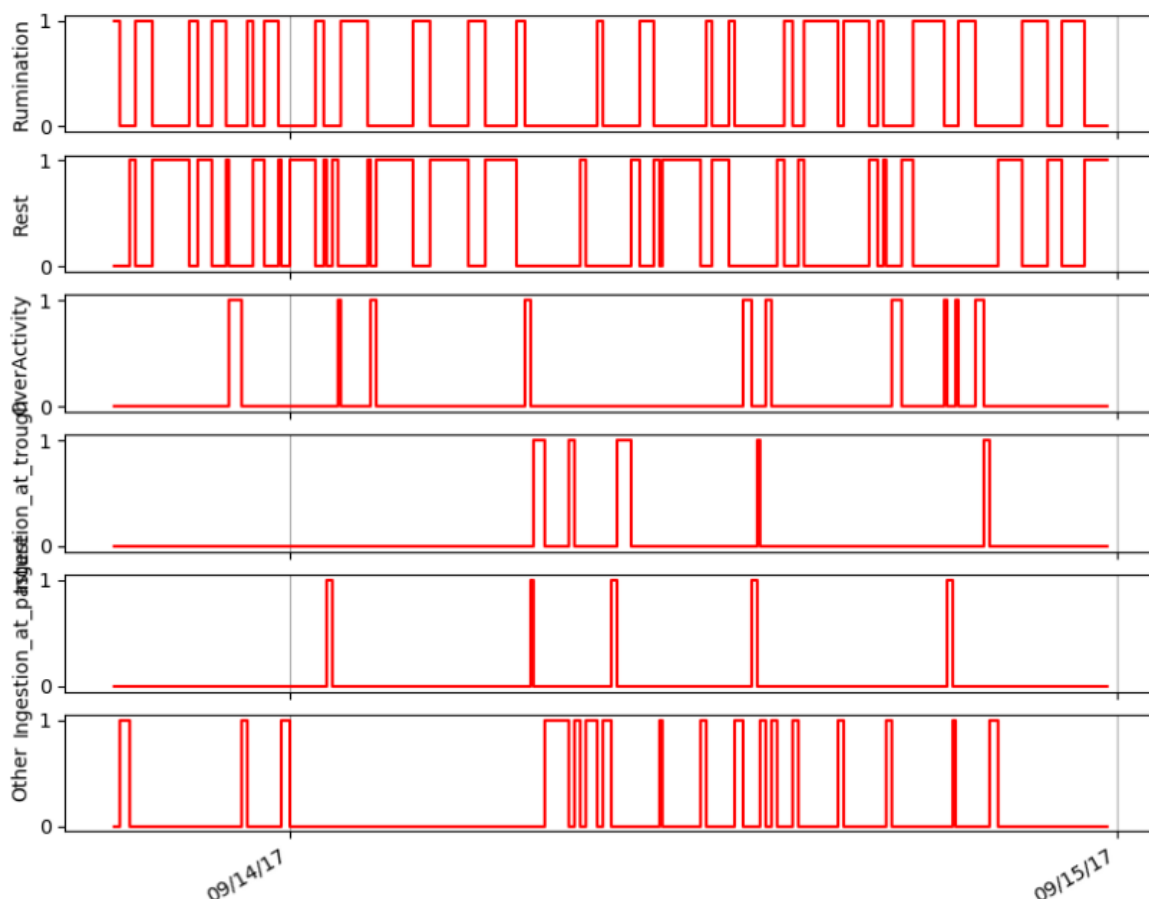
Other Activity: *True* si l'activité ne peut pas être déterminé

Une seule valeur peut être à *True* par pas de temps.

2. Classification du **debout/couché:**

Up: *True* si l'animal est debout, *False* s'il est couché.

Exemple de données normalisées:



Chaque fichier, nommé `normalized/<event_id>.parquet` correspond à une série temporelle associée à un événement présent dans les métas-données.

Ces fichiers sont sous le format parquet pouvant être lu directement avec la librairie pandas: `pd.read_parquet`.

6.2 Considérations légales

Ces données vous sont fournies dans le cadre du programme Agritech que vous suivez. Elles ne peuvent en aucun cas être diffusées, utilisées pour d'autres finalités que ce projet. Tout travail effectué sur ces données implique votre acceptation sans restriction de ces règles.

Au moment de lire ce sujet, vous devez avoir signé l'engagement de confidentialité individuel qui vous engage à ne pas divulguer toutes les informations que vous allez utiliser dans le cadre de ce projet. Si ce n'est pas le cas, contactez immédiatement votre responsable 42 pour le cursus Agritech et n'accédez pas aux données pour le moment.

7. Suivi

L'équipe de ITK effectue une présentation générale du projet, puis à intervalles réguliers se propose d'interagir avec les groupes d'étudiants travaillant sur le projet. Les créneaux d'interaction seront communiqués à l'avance par votre responsable Agritech.

8. Rendu et évaluation

Chaque partie fait l'objet d'un rendu dans un répertoire séparé, sur le repo git attribué par l'intra lors de l'inscription de l'équipe.

L'évaluation du projet sera dans la mesure du possible, conjointement effectuée par les équipes de ITK, un ou plusieurs exploitants, et un ou plusieurs pairs dans la promo Agritech.