

06/11/2018

Scope review

Projet inno : Mooveat



latitudes
exploring tech
for good_

mooveat
reconnectons-nous

Michel Richard, Othmane Jebbari, Alassane Diallo
CENTRALESUPELEC

Page 0 | 4

1. Contexte

Le projet de Mooveat est de produire une plateforme qui permette à l'agriculteur de gérer sa production de manière raisonnée, en faisant matcher offre et demande. Cela passe donc par 3 fonctionnalités : détermination de l'offre, détermination de la demande et simulation d'une potentielle nouvelle offre (installation ie Achat d'une nouvelle parcelle par exemple).

2. Reformulation de notre besoin

Notre projet s'inscrit dans les parties offre et installation. Le besoin est de simuler un rendement potentiel d'une certaine parcelle agricole aux vues d'informations données par l'agriculteur : variété plantée, surface considérée, localisation. Le but est que cette simulation ait une grande granularité avec une dépendance directe à la variété cultivée. Ainsi le but serait pour nous de le faire au niveau de 2 ou 3 produits assez éloignés (par exemple lait et gariguettes) et voir ce qui est transposable

Remarque : Des informations comme le prix unitaire permettront aussi de simuler la Demande et les revenus théoriques.

3. Objectifs

L'objectif fixé est d'avoir produit un back-end qui soit opérationnel. Si le planning nous le permet deux autres parties pourront être ajoutées au livrable : Front-End et/ou réflexion sur la manière de réutiliser le code produit pour les autres fonctionnalités décrites dans la fiche projet de départ (voir planning).

4. Phase d'exploration

Après les nombreuses séances passées à faire de la recherche documentaire pour à la fois comprendre le contexte du projet et ses notions clés et rechercher des données utiles à l'algorithme, nous avons trouvé 4 bases de données qui nous semblaient utiles pour l'algorithme.

Les 3 premières proviennent de la plateforme Agri4cast (open-source). Ce sont des données satellites avec un maillage de 25km sur 25km.

- Grids: (X_{min} , X_{max} , Y_{min} , Y_{max} , Grid code, Grid Id, hauteur). Les données qui nous semblent intéressantes sont les données de localisation (X et Y) et le Grid Id pour pouvoir localiser la parcelle et croiser les données avec les autres bases de données.
- Données météorologiques : (Grid_Id, Latitude, Longitude, Altitude, day T_{min} , T_{max} , T_{mean} , V_{vent} , P_{atm} , précipitations, radiations, snow depth). Ici, ce sont toutes les données météorologiques qui nous semblent utiles (T_{min} , T_{max} , T_{mean} , V_{vent} , P_{atm} , précipitations, radiations, snow depth). Il reste à déterminer quels paramètres sont vraiment essentiels dans la détermination des rendements.
- Gridded remote sensing data (Id_Grid, Sensor_Id, Day, Value, Id_cover). "Value" correspond à l'indice de végétation, c'est-à-dire la surface couverte par la végétation

sur la surface du grid. Nous pensions pouvoir remonter au rendement à partir de cette donnée éventuellement avec du traitement d'image (en supposant que la relation entre indice de végétation et rendement existe).

- Yearly modeled crop area (Id_grid, Id_crop, year area, ratio). Le type de produit agricole (blé, riz, betterave, etc.) produit sur le grid correspondant est ce qui pouvait nous servir. Si le rendement était déterminable à partir de l'indice de végétation, il aurait aussi dépendu du type de produit. Nous pensions donc qu'en croisant ces deux données nous pouvions peut-être déterminer le rendement des parcelles et ainsi avoir les outputs test pour entraîner notre modèle.

La dernière base de données est une base de données sur les sols.

- Gis sol : données sur plus de 26 millions d'échantillons de terre entre 2004 et 2009. On a les quantités chaque type de minéral dans les échantillons. Il nous reste à déterminer quels minéraux influent vraiment et voir s'il est possible de lier quantité de pesticides avec les minéraux présents dans le sol.

5. Planification

Tout d'abord, nous nous sommes rajoutés une séance de travail par semaine le jeudi (14h-17h) car nous sommes tous libres et de ne se rejoindre qu'une fois par semaine ne nous permettait pas de mettre en commun nos informations assez souvent et endiguait notre avancée.

Nous avons aussi décidé d'être disponibles le plus possible les samedis pour terminer les objectifs de la semaine si nous ne l'avons pas fait le jeudi soir.

De plus, nous avons aussi mis en commun l'ensemble de nos impératifs (qu'ils soient scolaires ou personnels), nous avons donc connaissance de l'ensemble des séances de travail qu'il reste à notre disposition. Nous avons aussi déterminé l'ensemble des deadlines et jalons liés au projet pour pouvoir adapter notre planning en conséquence.

Par ailleurs, nous avons décidé de « faire d'une pierre deux coups » et d'utiliser notre projet inno et l'algorithme que nous devons produire pour notre projet de Machine Learning. Outre l'avantage de pouvoir faire un deux en un, cela nous rajoute un deadline au 6/01 (10 jours avant notre soutenance finale) ce qui nous permettra de nous concentrer ensuite sur des potentielles améliorations ou encore le Front-End.

Pour ce qui est de la formation en Machine Learning, nous la faisons en parallèle sur notre temps libre.

Voici le planning que nous avons établi :

- 06-nov Définir ce qu'on fait une fois le code opérationnel
Définition planning
Définir étapes de l'algorithme
Avoir l'ensemble des données dont on a besoin à disposition avant la fin de la semaine
Mise en forme du Jalon 2
- Avoir un modèle précis de l'algo (étapes, input/output, modèle associé), Mise en forme du
- 13-nov jalon 2
- 20-nov Forum CentraleSupélec
- 22-nov Mise en forme Jalon 2 et début du codage
- 27-nov Jalon 2
Explication claire des objectifs
Résumé de la recherche
Explication de l'algo
Mise en forme
- 28-nov Exam Electif 2
- 29-nov Exam Eco
- 02-déc Assignement 2 : Machine learning
- 07-déc Exam Electif 3
- 14-déc Exam d'Alassane
- 18-déc Mi-projet : point sur ce qu'on a déjà fait
- Vacances de Noel (possibilité de travailler sur la partie Front-end ou généraliser le modèle
- 24-déc utilisé sur les deux autres algos)
- 06-janv Deadline pour rendre le rapport et le code du projet en tant que projet ML

Nous avons décidé qu'à partir de la semaine où nous commençons le codage nous travaillerons sur des objectifs semaine à semaine, chaque semaine étant terminée par un debrief et une remise en forme des objectifs et éventuellement du modèle.

Pour les deux dernières semaines nous comptons réfléchir aux différentes ouvertures possibles et préparer la soutenance finale.

6. Risques

Nous voyons principalement deux risques à ce projet et notre manière de l'avoir planifié. Le premier est qu'il n'y ait pas de Machine Learning dans le projet. En effet, cela représenterait une grande perte

de temps pour nous et rendre beaucoup plus difficile de rendre le livrable car nous aurions quand même le projet de machine learning en parallèle.

L'autre risque que nous avons identifié est que nous nous perdions dans trop de données. Nous ressentons donc le besoin de vraiment cadrer plus le besoin pour pouvoir se recentrer sur la recherche des données nécessaires.

7. Ce que l'on a retenu

- Deux approches possibles :
 - Soit l'approche Machine Learning qui nécessite une base de données sur les rendements passés, que ce soit nous qui la construisions ou que l'on trouve les données brutes, elles sont nécessaires pour entraîner l'algorithme.
 - Soit l'approche « classique » : trouver l'influence de chacun des paramètres que l'on considère. De manière grossière, trouver un moyen de faire le travail à la place de l'algorithme. Dans tous les cas il sera intéressant de faire des recherches dessus car ça pourrait nous permettre d'éliminer/rajouter/modéliser certains paramètres.

Nous ce qui nous arrangerait serait d'utiliser l'approche Machine Learning, c'est ce qui nous intéresse le plus vis-à-vis notre formation mais aussi de la faisabilité du livrable.

- En discutant avec Thibault nous nous sommes rendus compte que nous ne pourrions pas trouver une relation entre indice de végétation et rendement, il nous faut donc trouver un autre moyen d'y parvenir.
- Par ailleurs, le découpage en grid n'est pas adapté, on a besoin d'un découpage en parcelles. On a donc besoin de croiser la base de données d'Agri4Cast avec celle d'API-Agro qui recense les parcelles. Utiliser le logiciel Agromonitoring nous permettrait d'avoir cette étape déjà faite (le logiciel donne les données météorologiques de chaque parcelle au cours du temps) mais il est payant.
- Il y a une grosse différence entre le fonctionnement pour les produits issus directement de l'agriculture et ceux issus de l'élevage. Pour l'instant nous nous sommes plus penchés sur le côté agricole, il faudrait aussi penser à se pencher sur l'élevage.
- Impossibilité de faire du crowdsourcing auprès des agriculteurs
- Difficulté pour trouver des données sur les pesticides ou encore la fréquence d'irrigation car ce sont des données que les agriculteurs ne sont pas enclins à partager.
- Contacter Céline pour essayer d'être mis en contact la startup CybelTech fondée par des Centraliens.