

Tirer avantage de données auxiliaires et d'interactions utilisateurs afin d'apprendre avec peu de données.

Leveraging user-interaction and auxiliary data to learn from small data.

Romain Mormont

Unité systèmes et modélisation,
Département d'électricité, électronique et informatique,
Université de Liège, Belgique

27 septembre 2016

Contexte

- Le **machine learning** (ML) a récemment connu plusieurs succès commerciaux et industriels (DeepMind, voitures auto-pilotées Google et Tesla, IBM Watson,...)
- Ces succès sont souvent dûs à deux facteurs :
 - **Abondance de données** (*big data*)
 - **Grande puissance de calcul**
- Ces **critères sont essentiels pour appliquer la plupart des méthodes de ML**

Problème

- **Ces critères ne sont pas toujours remplis.**
- Les données disponibles ne le sont pas toujours quantité suffisantes. On peut alors parler de **small data** !

Small data : "la quantité de données n'est pas grande assez relativement à la tâche à résoudre sans pour autant être petite de manière absolue"

Applications *small data*

Est-ce que je présente des applications ici, où j'attends le case study ?

Objectifs

Explorer et développer de nouvelles méthodes applicables aux problèmes *small data*.

Pistes et questions de recherche :

1. Comment **intégrer l'humain aux procédures d'apprentissage** afin d'en améliorer les performances ?
2. Comment **exploiter des données auxiliaires** disponibles afin d'atteindre le même objectif ?

Question 1 : humain dans la boucle ?

TODO : Illustration of the "human-in-the-loop" with diagrams ?
(one with ML expert, one with domaine expert interacting with algo ?)

Question 2 : données auxiliaires ?

TODO : Again with an illustration ? Transfert learning, imprecise/precise data, multi-modal images,...

Cas d'étude

Illustration cytomine

⇒ rare object detection and categorization in high-resolution tissue images

⇒ en production peut être utilisé pour collecter du feedback utilisateur

Autres applications (?)

Détection de défaut image de pièces ? ?

Pepite ? ?

Merci pour votre attention !
Des questions ?

Backup slides

TODO :

- calendrier,
- résultats du TFE,
- applications à moyen ou long terme,
- publications récentes. . .
- théorie (ML,...)