

WEBINAIRE

**GESTION DE LA CONFIGURATION ET
DES RÉSULTATS AVEC MLFLOW,
HYDRA ET POUTYNE**

30 OCTOBRE 2020

OBJECTIFS DE LA PRÉSENTATION

- Initier aux outils de gestion de l'entraînement, de la configuration et des résultats.
- Développer de bonnes pratiques.
- Améliorer votre productivité.


VOTRE CONFÉRENCIER



DAVID BEAUCHEMIN

Candidat au doctorat

Département d'informatique et de génie logiciel

- Introduit à la recherche reproductible en 2016 (R Markdown et **git**)
- Participation à REPROLANG de la conférence LREC [Garneau et al., 2020]
- Membre actif dans le développement d'une librairie facilitant la reproductibilité (**Poutyne** *)

AU MENU



Gestion de la configuration



Gestion des résultats

La gestion d'un projet

GESTION DES PARAMÈTRES DE CONFIGURATION

```
@experiment.config
def config() :
    seed = 42
    num_runs = 10
    iteration = 0
    source_language = "en"
    target_language = "de"
    src_input = "path" # The input source embeddings
    trg_input = "2e path" # The input target embeddings
    other_input = "3e path" # Commentaire pas clair
    :
    (ligne 500) ne paramètre
```

GESTION DES PARAMÈTRES DE CONFIGURATION

Quel paramètre fait
ça déjà ?

GESTION DES PARAMÈTRES DE CONFIGURATION

Quel paramètre fait
ça déjà ?

Quels paramètres
sont nécessairement
ensemble ?

GESTION DES PARAMÈTRES DE CONFIGURATION

Quel paramètre fait
ça déjà ?

Quels paramètres
sont nécessairement
ensemble ?

Sont-ils tous
essentiels ?

GESTION DES PARAMÈTRES DE CONFIGURATION

Quel paramètre fait
ça déjà ?

Quels paramètres
sont nécessairement
ensemble ?

Sont-ils tous
essentiels ?

Comment doit-on se
retrouver dans le
projet ?

GESTION DES RÉSULTATS

- res_1.txt
- res_2.txt
- res_3.txt
- res_4.txt
- res_5_good.txt
- res_5.txt
- res_6_fix_a.txt
- ⋮
- n^{e} fichier de résultats

GESTION DES RÉSULTATS

Quelle configuration
déjà avec ces
résultats ?

GESTION DES RÉSULTATS

Quelle configuration déjà avec ces résultats ?	Est-ce un succès ou un échec cet entrainement ?
--	---

GESTION DES RÉSULTATS

Quelle configuration
déjà avec ces
résultats ?

Est-ce un succès ou
un échec cet
entraînement ?

Lequel contient mes
meilleurs résultats ?

GESTION DES RÉSULTATS

Quelle configuration
déjà avec ces
résultats ?

Est-ce un succès ou
un échec cet
entraînement ?

Lequel contient mes
meilleurs résultats ?

Comment doit-on se
retrouver dans les
résultats ?

Les solutions

AU MENU



Gestion de la configuration



Gestion des résultats

GESTION DE LA CONFIGURATION



Simple et efficace

GESTION DE LA CONFIGURATION



Simple et efficace



Facilite l'expérimentation

GESTION DE LA CONFIGURATION



Simple et efficace



Facilite l'expérimentation



Extensible



Open source et
licence MIT



Fichiers de
configurations
structurés YAML



Fichiers de
configurations
hiérarchiques



Balayage de
configurations

CONFIGURATION STRUCTURÉ

```
conf
├── config.yaml
├── dataset
│   ├── canadian.yaml
│   └── netherlands.yaml
├── embeddings
│   └── fast_text.yaml
├── model
│   ├── bi_lstm_bidirectionnal.yaml
│   ├── bi_lstm.yaml
│   ├── lstm_bidirectionnal.yaml
│   └── lstm.yaml
├── optimizer
│   ├── adam.yaml
│   └── SGD.yaml
```

CONFIGURATION HIÉRARCHIQUE

data_loader :

batch_size : 2048

setting :

seed : 42

device : "cuda:0"

defaults :

- optimizer : SGD

- model : bi_lstm

- dataset : canadian

- embeddings : fast_text

trainer :

num_epochs : 1

patience : 30

CONFIGURATION HIÉRARCHIQUE

optimizer : SGD

```
optimizer :  
  lr : 0.1  
  type : sgd
```


BALAYAGE DE CONFIGURATIONS

```
python main.py -multirun task=1,2,3,4,5
```

```
python main.py -m 'main.x=int(interval(-5, 5))' 'main.y=int(interval(-5, 10))'
```

EN BONUS

- Journalisation automatique et personnalisable
- Instanciation paramétrique

```
# @package _group_  
_target_ : my_app.MySQLConnection  
host : localhost  
user : root  
password : 1234
```

POINT NÉGATIF

```
hydra.utils.get_original_cwd()
```

AU MENU



Gestion de la configuration



Gestion des résultats

GESTION DES RÉSULTATS



Simple à utiliser

GESTION DES RÉSULTATS



Simple à utiliser



Journalisation des
expérimentations

GESTION DES RÉSULTATS



Simple à utiliser



Journalisation des
expérimentations



Visualisation rapide des
expérimentations

MLFLOW TRACKING *



Open source et
licence Apache 2.0



Suivie automatique
de paramètres
d'entraînement



Visualisation simple



Intégration avec
Poutyne

SUIVIE AUTOMATIQUE DE PARAMÈTRES D'ENTRAÎNEMENT

- Version du code (**git**)*
- Horodatage de l'entraînement
- Succès/échec de l'entraînement
- Configuration de l'ordinateur
- Utilisateur

VISUALISATION SIMPLE

```
mlflow server -p 5000 -h 127.0.0.1 --backend-store-uri file:///absolute/path
```

VISUALISATION SIMPLE

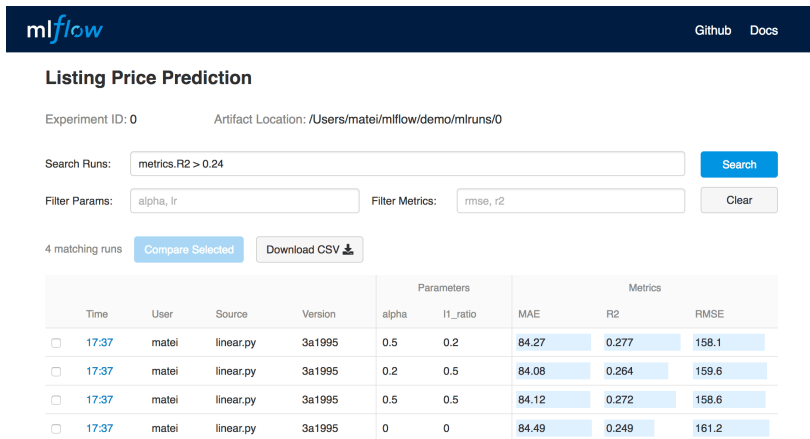


Figure 1 – Introducing MLflow : an Open Source Machine Learning Platform *

VISUALISATION SIMPLE

- Tri sur les expérimentations
- Recherche des expérimentations
- Requêtes sur les résultats
- Exportation des résultats
- Visualisation les métriques

INTÉGRATION AVEC **POUTYNE** *

La version de « base » implique de

- journaliser manuellement les paramètres de configuration,
- journaliser manuellement les métriques à chaque étape et itération,
- journaliser manuellement la version du code.

INTÉGRATION AVEC POUTYNE *

La solution, [MLFlowWriter](#), un *callback* permettant de

- journaliser semi-automatiquement les paramètres de configuration,
- journaliser automatiquement les métriques à chaque étape et itération,
- journaliser automatiquement la version du code,
- journaliser manuellement un modèle,
- journaliser automatiquement les métriques de test lors d'une phase de test.

POINT NÉGATIF

La documentation n'est pas toujours facile à naviguer.

La suite

PRÉSENTATION DES RÉSULTATS



Génération automatique des tableaux




Rapport dynamique



Itérations d'expérimentations

POUR ALLER PLUS LOIN (EN ORDRE)

- Notification de l'état d'entraînement **Notif** *
- **Continuous Machine Learning (CML)** *

PÉRIODE DE QUESTIONS




WEBINAIRE

**MERCI DE VOTRE
ÉCOUTE!**



REFERENCES i

-  Garneau, N., Godbout, M., Beauchemin, D., Durand, A., and Lamontagne, L. (2020).
A Robust Self-Learning Method for Fully Unsupervised Cross-Lingual Mappings of Word Embeddings : Making the Method Robustly Reproducible as Well.