



REDS MÉTHODOLOGIE RECHERCHE

Expérimentation

Thursday 21st December, 2023

Laure Soulier

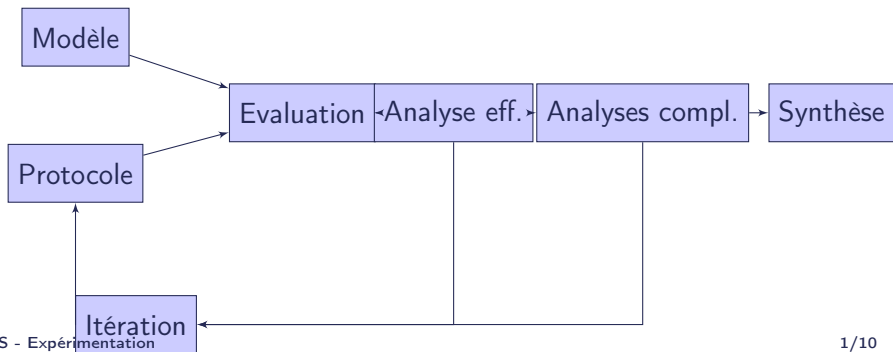


Définition générale

Définition générale

Objectifs

- Tester la validité d'un modèle
- Comparaison de modèles
- Paramétrage d'un modèle
- Comprendre le fonctionnement d'un modèle



Protocole d'évaluation

Protocole d'évaluation

- Jeux de données (description + statistiques)
 - Génériques : Kaggle, Google Dataset Search,
 - <https://pub.towardsai.net/best-datasets-for-machine-learning-data-science-computer-vision>
 - https://en.wikipedia.org/wiki/List_of_datasets_for_machine-learning_research
 - Derniers datasets: <https://twitter.com/paperswithdata>
 - NLP: <https://index.quantumstat.com/>
 - Vision: <https://public.roboflow.com/>
 - Possibilité de générer son propre dataset à partir d'autres datasets

Choix des datasets

On ne fait pas de la recherche car on a un dataset, mais la recherche est guidée par des hypothèses. Le dataset est utilisé/construit à partir des hypothèses de recherche

Protocole d'évaluation

- Jeux de données (description + statistiques)
- Métriques
 - Mesures d'évaluation automatiques (erreur, MAP, precision/rappel)
 - On pourrait aussi avoir des mesures plus "subjectives" / liées à l'humain (évaluation utilisateur ?)
 - Significativité des métriques !!! tests, variances,

Choix des métriques

- Métriques standards sur la tâche
- Métriques en fonction des hypothèses à tester

Attention à ne pas biaiser son choix à partir de ce qu'on veut obtenir.

Protocole d'évaluation

- Jeux de données (description + statistiques)
- Métriques (erreur, MAP, précision/rappel)
- Modèles de référence (nom + objectif de comparaison)
 - Résultats issus des papiers
 - Implémentation des modèles disponible (github/gitlab, ...). On lance juste le modèle.
 - Réimplémentation (prouver que votre ré-implémentation est ok).

Choix des baselines

- Listing des modèles SOTA (surtout en vision)
- Choix en fonction des hypothèses à tester

Protocole d'évaluation

- Jeux de données (description + statistiques)
- Métriques (erreur, MAP, precision/rappel)
- Modèles de référence (nom + objectif de comparaison)
- Liste des caractéristiques (*features*)
 - suivant les modèles, on peut être amené à définir des caractéristiques. Bien les définir (nom, intuition, formule)

Protocole d'évaluation

- Jeux de données (description + statistiques)
- Métriques (erreur, MAP, precision/rappel)
- Modèles de référence (nom + objectif de comparaison)
- Liste des caractéristiques (*features*)
- Vérité de terrain (*ground truth*)
 - Fournie dans le jeu de données. Sert à calculer les métriques
 - Ce n'est pas forcément l'objectif ultime... surtout pour des tâches de génération...
 - Utiliser autant que possible le jeu de données tel que utilisé dans d'autres papiers pour avoir des comparaisons rapides entre modèles.

Protocole d'évaluation

- Jeux de données (description + statistiques)
- Métriques (erreur, MAP, précision/rappel)
- Modèles de référence (nom + objectif de comparaison)
- Liste des caractéristiques (*features*)
- Vérité de terrain (*ground truth*)
- Méthodologie (nettoyage des données, cross-validation, détails sur l'architecture d'un réseau...)
 - Utile pour la reproductibilité des données
 - Permet de juger de la fiabilité de votre protocole. Si mal fait, on peut arrêter de lire les résultats et ne pas y croire.

→ Voir le document de synthèse

→ Papiers CIKM (role factor), CIKM (answering Twitter), ICML (ZSL), ECIR (CF)

Analyses

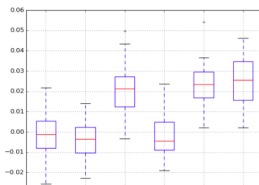
- Analyse d'efficacité : comparaison avec les modèles de référence

Model	NDCG@20	% Change
BM25	0.429	+59.77%***
Hiemstra	0.322	+113.13%***
PRank	0.641	+7.03%*
BibRank	0.686	

Model	NDCG@20	% change
BM25	0.376	+38.26%***
Hiemstra	0.428	+21.47%**
PRank	0.455	+14.29%*
BibRank	0.520	

Figure 1: [Soulier et al., 2012]

- Analyse complémentaire ou *failure analysis* : par requête / par dataset / par classe / par intervalle de temps, ...



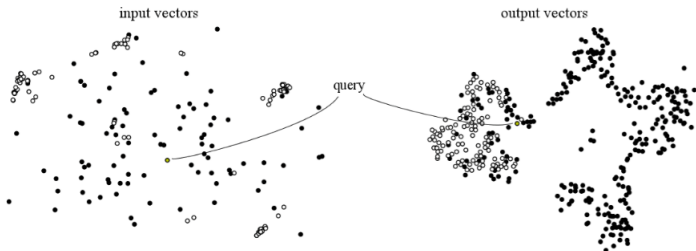
Analyses

■ Analyse des facteurs

Feature	Sandy	Ebola
	Regres. estimate	Regres. estimate
Topical	-	-0.23***
Images	-	-0.19***
Links	-0.08***	-
Hashtags	-0.14***	1.58***
Only text	-0.08***	-0.90***
Positive opinion	-0.12***	-1.33**
Neutral opinion	-0.16***	-
Negative opinion	-0.16***	-

Figure 3: [Tamine et al., 2016]

■ Projection visuelles



Les astuces et les pièges à éviter

- Minimiser l'effort du lecteur : légende lisible, titre, axes compréhensibles, échelles, unités, légendes explicative...
- Simplifier les graphiques : pas trop d'information, pas trop d'échelles différentes
- Comparer des figures à la même échelle et même origine
- Figure référencée dans le texte