

TD 3 - Expériences et données

Carlos Ramisch et Manon Scholivet - Recherche Zen

2023-2024

L'objectif de cette activité est de vous familiariser par la pratique avec certains outils de la préparation d'expériences, notamment la création et évaluation de jeux de données, ainsi que la conception du protocole et des mesures d'évaluation. Les activités ci-dessous sont relativement indépendantes les unes des autres (sauf les deux premières) et seront effectuées tout au long du cours.

1 Annotation de données (15 min)

L'annotation de données pose bien de nombreux problèmes qu'il est parfois difficile d'appréhender sans les avoir expérimentés. Nous vous proposons de vous familiariser avec une tâche d'annotation qui concerne des expressions polylexicales (EPL). Une EPL est un groupe d'unités lexicales qui doivent être considérées comme une unique unité lexicale. Par exemple : "C'est la *douche froide*." Ici, on vous demande d'annoter à quel point chaque expression est "compositionnelle", c'est-à-dire, à quel point le sens de l'expression peut être déduit à partir des sens des mots qui composent l'expression pris séparément. Vous verrez qu'une dizaine d'annotations seulement peuvent déjà poser bien des soucis !

Rendez-vous sur le lien : https://parsemefr.lis-lab.fr/compounds_fr/.

Et suivez les instructions !

2 Accord inter-annotateur.ice (10 min de débrief)

Après avoir annoté une dizaine d'expressions, vous devriez être plus à l'aise avec cette tâche. Mais avez-vous annoté de façon cohérente avec le reste du groupe ?

Pour répondre à cette question, nous pouvons calculer vos scores d'accord inter-annotateur.ice! Nous allons voir ce qu'il est possible de conclure de ces scores.

3 Métriques d'évaluation (25 min + 20 min débrief)

Après avoir annoté des EPL vient l'étape de leur prédiction dans un nouveau texte ! Mais comment vérifier la qualité de la prédiction ?

Il va être nécessaire de créer une mesure d'évaluation. Mais vous allez rapidement être confrontés à un certain nombre de problèmes. En voici quelques uns :

- Les EPL peuvent être discontinues ! Par exemple : "Il *prend* ces problèmes *en compte*."
- Que faire si seulement une partie de l'EPL est prédite ?
- Que faut-il évaluer ? La précision ? Le rappel ? Autre chose ?
- Faut-il accorder plus de valeur à une EPL plus difficile à détecter, par exemple qui n'a pas été vue dans le corpus d'entraînement ?

Pour vous guider, vous pouvez utiliser un exemple comme celui-ci, qui illustre certains des défis listés ci-dessus :

Gold: *make* segmentation *decisions*, *in order to* split sentences into *lexical units*

System: *make* segmentation *decisions*, *in order to* *split sentences* into lexical *units*

À vous de trouver la (ou les ?) meilleure(s) métrique(s) ! Essayer de formuler la métrique de manière assez précise, voire la formaliser avec des équations si vous y arrivez.

4 Protocole expérimental (*30 min + 20 min debrief*)

Vous rappelez vous de votre question de recherche définie lors du premier cours ? C'est le moment de la ressortir, et d'établir un protocole expérimental qui vous permettra d'y répondre !