



IFT6285 (TALN) — Projet 1
Classification

Contact :
Philippe Langlais +1 514 343 61 11 ext: 47494
RALI/DIRO felipe@iro.umontreal.ca
Université de Montréal <http://www.iro.umontreal.ca/~felipe/>

■ dernière compilation : 11 octobre 2021 (23:51)

Contexte

Avec les progrès en IA ont émergés des jeux de données comme [GLUE](#) ou [SuperGLUE](#) qui contiennent chacun une dizaine de tâches à saveur sémantique. Ces jeux de données sont réunis dans le but de favoriser une approche de transfert où un modèle pré-entraîné est ajusté sur une ou plusieurs tâches en même temps. De ce fait, un modèle entraîné directement sur une tâche à très peu de chance de battre une approche de transfert.

Vous n’avez cependant aucune obligation de considérer dans ce projet une approche de transfert.

Données

Voici quatre tâches réunies dans [GLUE](#) qui sont toutes des tâches de classification binaire :

COLA [The Corpus of Linguistic Acceptability](#): consiste à identifier si une phrase est acceptable ou pas.

SST [The Stanford Sentiment Treebank](#) où la tâche consiste à identifier si la phrase d’entrée est positive ou pas

QQP Quora Question Pairs, consiste à identifier si deux questions sont des paraphrases

MRPC [Microsoft Research Paraphrase Corpus](#), consiste à identifier si deux phrases sont des paraphrases

Vous allez considérer (au moins) deux de ces quatre tâches en vous assurant de considérer COLA ou SST (entrée = une phrase) ainsi que QQP ou MRPC (entrée = deux phrases). Si vous remettez vos travaux en groupe de trois personnes, alors au moins trois de ces tâches doivent être considérées.

Pour la plupart des tâches, le corpus de test est réservé pour les soumissions officielles (vous n’avez pas la réponse). Vous évalueriez donc vos approches sur le corpus de développement de chaque tâche.

À faire

1. Téléchargez les données des quatre tâches. Analysez les.
2. Entraînez un système de base (*baseline*) par tâche et indiquez sa performance sur le corpus de développement associé à la tâche. Vous prendrez soin de décrire votre baseline ainsi que la ou les métriques d'évaluation utilisées.
3. Décrire vos efforts pour créer un modèle qui améliore les performances obtenues par votre baseline. Vous identifierez pour le ou les modèles proposés les hyper-paramètres qui vous permettent de les ajuster. Vous montrerez l'impact de ces hyper-paramètres (ou un sous-ensemble).
4. Identifiez la configuration de votre meilleur système pour chaque tâche étudiée (la même configuration peut être utilisée pour toutes les tâches) ainsi que sa performance.
5. Analyser finement les performances de votre meilleur modèle en créant à cet effet de nouvelles données dans le but de mieux le comprendre. Vous décrirez le résultat de cette analyse. Pour chaque tâche, vous devez créer au moins 50 nouvelles données de test que vous sauvez dans un fichier du même format que le jeu de développement accompagnant chaque tâche et portant le nom `tâche-nom.t/csv` où `tâche` $\in \{\text{qqp}, \text{mrpc}, \text{cola}, \text{sst}\}$ et `nom` indique l'identité des personnes impliquées dans le projet. Vos données ne devront pas être toutes de la même classe.
6. Si l'analyse de votre meilleur modèle vous aide à identifier une lacune que vous êtes en mesure de corriger, faites le et indiquez ce que cela donne.
7. Une fois votre projet remis (après le 12 novembre), un jeu de test sera distribué pour chaque tâche. Vous lancerez votre meilleur modèle pour les tâches que vous avez étudiées. Le format de vos sorties doit être au même format que celui des fichiers `dev` de chaque tâche. Plus après la remise.

Le rapport (format pdf, en français ou en anglais) décrivant vos expériences doit faire au plus cinq pages et doit porter le nom `rapport-nom.pdf` où `nom` indique votre identité.

Notation

La notation n'est pas corrélée à la performance de vos approches, mais à la **curiosité** que vous développerez et à votre esprit d'**analyse**. La clarté et l'informativité de vos rapports sont deux critères importants.

Voici quelques questions que nous poserons lors de la correction :

- les modèles baselines sont-ils raisonnables ?
- les efforts mis en place pour dépasser les performances du modèle de base sont-ils justifiés, bien décrits, conséquents ?
- le rapport est-il clair, les analyses sont-elles informatives ?
- les ressources annotées sont-elles faites avec soin ?

Dans la mesure où ces jeux de données sont très étudiés, il existe de nombreux codes que vous pouvez utiliser. Le faire n'est aucunement un problème, mais vous devez indiquer très clairement les codes que utilisez.

Remise

La remise est à faire sur Studium sous le libellé **projet1**. Vous devez remettre votre code, votre rapport (format pdf, texte en anglais ou en français), ainsi que les fichiers d'annotation (un par tâche étudiée) dans une archive (gzip, tar, tar.gz) dont le nom est préfixé de **projet1-noms**, où **noms** est à remplacer par l'identité des personnes (**prénom_nom**) impliquées dans le projet. Donc si j'avais à remettre seul mon solutionnaire au projet1, je le ferais sous le nom **projet1-philippe_langlais.tar.gz**. Assurez vous que le nom des personnes impliquées dans le projet soit indiqué sur le rapport. Le projet est à remettre en groupe d'au plus trois personnes au plus tard vendredi 12 novembre à 23h59.