

IAAA / PSTALN : Examen

Benoit Favre

28 janvier 2020

Document : supports et notebooks associés aux cours/tp autorisés. Pas de connexion internet !

Exercice 1 (4 points)

- Quels sont les désavantages des représentations 1-hot des mots ?
- Quelles sont les différences entre les CRF et les HMM (en terme d’hypothèses, de phénomènes modélisés, et de type de modèle) ?
- Pourquoi les modèles n-grammes entraînés sur de très gros corpus occupent-ils énormément de mémoire ?
- À quoi sert l’algorithme “forward” dans les CRF ?

Exercice 2 (3 points)

Pourquoi les mécanismes d’attention sont-ils de plus en plus au cœur des nouvelles architectures de traitement du langage naturel ?

Exercice 3 (5 points)

Selon vous, l’expertise linguistique est-elle encore nécessaire en traitement du langage naturel ? Argumentez.

Exercice 4 (8 points)

Vous êtes appelé à travailler pour une entreprise de commerce électronique qui vend des produits de beauté sur son site. Les produits sont associés à des notes et des avis donnés par les clients de cette entreprise. Les produits qui ont beaucoup d’avis se vendent relativement bien (si ces avis sont bons), mais ceux qui viennent d’être ajoutés au catalogue ne se vendent pas très bien. Suggérant l’hypothèse que c’est probablement le manque d’avis clients qui empêche les ventes des produits concernés de décoller, le gérant de l’entreprise vous charge de mettre au point une technique de génération pour créer de toutes pièces des avis plausibles.

Discutez les approches possibles pour résoudre ce problème, la méthodologie que vous mettriez en place pour vous assurer de la qualité du système produit, les problèmes de généralisation que vous risquez de rencontrer et comment vous pourriez les contourner. Vous en profiterez pour aborder les considérations éthiques liées à ce problème.

Votre réponse devra faire 1 page maximum. Lorsque le scénario n’est pas suffisamment précis, explicitiez les hypothèses que vous prenez.