

Classification de répliques dans Star Wars: droïde ?

Des robots à notre image, mais à quel point...

... à la recherche de la distance à laquelle on *choisit* de placer les robots de l'humain

Dans le cadre de ce projet:

- ▶ 424 répliques étiquetées à la main sur 5 films de la saga
- ▶ répliques en général brèves, voire très brèves
- ▶ répliques rares, et de personnages au vocabulaire très différent (fréquences de termes ☹): "Jedi scum" (Grievous) v. "My goodness!" (C-3PO)

Idéal:

Généraliser sur les supports (*I, Robot*; *Azimov*) et les formats (répliques brèves et rares v. longues et fréquentes)



Features, classifieurs et premiers résultats

Features, classifieurs et premiers résultats

Générer deux matrices TF-IDF:

1. lemmatisation (nltk WordNetLemmatizer)
2. étiquetage morphosyntaxique (nltk pos_tag)

	replica	isDroid	lemmatizedSentences	posTaggedSentences
0	I'm TC-14 at your service. This way, please.	1	I m TC14 at your service This way please	PRON VERB ADJ ADP PRON NOUN DET NOUN NOUN
1	We are greatly honored by your visit, Ambassad...	1	We be greatly honor by your visit Ambassadors...	PRON VERB ADV VERB ADP PRON NOUN NOUN NOUN V...

Classifieurs scikit-learn:

- ▶ Random Forest
- ▶ LinearSVC
- ▶ MultinomialNB
- ▶ LogisticRegression

Résultats: chiffres

Précision $\in [0.60; 0.66]$

Résultats: features

- parmi les plus discriminants pour la classe droide: sir,roger,r2
- moins informatifs pour le reste (non stéréotypé)