Achtatou Sean
Siwen GUO

Université du Luxembourg

Bachelor en Science Informatique

Le Machine Learning

pour l'analyse de sentiment

Semestre 3 - 2018-2019

1.Introduction

Ce projet de Bachelier fut réalisé par Sean Achtatou ; étudiant de la section en bachelier informatique de l'Université du Luxembourg étant actuellement diriger par le Prof. Dr. Nicolas Guelfi ; et supervisé par le PAT Siwen Guo.

Le projet met en avant la possibilité d'utilisation de l'Intelligence Artificielle pour la résolution de problèmes pouvant s'avérer long et peu efficace si réalisé par un humain. L'un des buts principaux est d'égailler la pensée humaine sur une possible utilisation d'Intelligence artificielle dans un futur proche ainsi que leur limite d'utilisation. Nous avons pu, par la suite, produire une l'Intelligence Artificielle permettant la classification de messages selon leur sentiment en utilisant le langage de programmation Python 3, étant de plus en plus utilisé pour l'Intelligence Artificielle.

De ce fait, nous avons utilisé une sous partie de l'Intelligence artificielle étant le Machine Learning, pouvant être utiliser pour pouvoir classifier un ensemble de données basés sur l'analyse et la prédiction de ces dernières.
Le type de Machine Learning utilisé fut supervisé, signifiant que les données utiliser étaient déjà classifiées. Le Machine Learning supervisé fournit également l'utilisation de différentes techniques de classification, nous avons choisi d'utilisé la machine de support de vecteurs qui sa base sur l'utilisation des données pour créer des vecteurs qui peuvent être utiliser pour développer des modèles de prédiction.

Le support vecteur machine se base sur plusieurs étapes :

- Prétraitement de données (modifie les données de sorte à être utilisé par la machine)
- Création des vecteurs d'entrainement et de test (basés sur les données)
- Insérer les vecteurs d'entrainement dans le modèle (linéaire ou kernel)
- Prédire la précision du modèle et son pouvoir de prédiction.

2.Sommaire

Nous avons eu l'occasion de pouvoir créer quatre diffèrent modèles.

Les deux premiers modèles se basaient sur un ensemble d'un millier de messages et devaient prédire leur sentiment étant positif ou négatif. Le deuxième modèle par rapport au premier, utilisa une technique différente étant le kernel trick, permettant d'utiliser un modèle en trois dimensions pour séparer les différents vecteurs entres eux. Nous avons pu obtenir les résultats suivants .**[Image1]**

Les deux derniers modèles se sont basés sur un ensemble d'un million de messages, mais seulement concernant ceux ayant une possibilité de lien avec un comportement terroriste. Pour cela nous avons due trier les messages par rapport à plusieurs mots reliés à un possible comportement terroriste. De plus, les deux modèles se sont basés sur l'utilisation de la libraire Glove, permettant d'obtenir des vecteurs de données de longueur fixe et de moindre longueur permettant une meilleure précision et une plus grande rapidité de création de modèle et de vecteurs. Également, le second modèle par rapport au premier utilisa le kernel trick pour son modèle. Nous avons pu obtenir les résultats suivants.**[Image2]**

Nous avons pu constater, grâce à l'observation résultant de nos différents résultats, que la plupart de nos modèles peuvent être potentiellement utilisable, comme classificateur de sentiment de messages dans une situation réelle.

3.Conclusion

Nous avons pu conclure de ce projet, que nos modèles utilisant le Machine Learning peuvent être utiliser comme de bon classificateur de sentiment. Nous avons pu également observer que la librairie Glove permettait une meilleure prédiction vis-à-vis de nos modèles lorsqu'il fut utilisé. Cela dit, nos modèles ne prennent pas en compte une situation concernant une possible ironie venant des messages, pouvant donc mener à une mauvaise prédiction de la part des modèles sur le sentiment.

Appendix

Prédiction/Sentiment	Positif	Négatif	Précision du modèle
Premier modèle (lineaire)			86 %
Prédiction correcte	93 %	89 %	
Prédiction incorrecte	7 %	11 %	
Second modèle (kernel trick)			82 %
Prédiction correcte	96 %	98 %	
Prédiction incorrecte	4 %	2 %	

Image 1. Résultats des deux premiers modèles pour le premier corpus sans utilisation de la librairie Glove.

Prédiction/Sentiment	Positif	Négatif	Précision du modèle
Premier modèle (linear)			61 %
Prédiction correcte	89 %	72 %	
Prédiction incorrecte	11 %	28 %	
Second modèle (kernel trick)			64 %
Prédiction correcte	93 %	96 %	
Prédiction incorrecte	7 %	4 %	

Image 2. Résultats des deux derniers modèles pour le second corpus avec utilisation de la librairie Glove.