

Influenceurs sur Twitter

Mots-clé : Influencer, marketing, Twitter, Neo4J

Auteur : PHELIZOT Yvan

Description du contexte

Les réseaux sociaux sont apparus comme les nouveaux médias de prédilection pour la nouvelle génération. Celle-ci y trouve en général sa source d'information principale, souvent devant des médias plus traditionnelles comme la télévision ou les journaux. Plus récemment encore, une nouvelle catégorie de personnes, les influenceurs ¹, sont devenus les nouveaux ambassadeurs des marques ou des partis politiques. Ils définissent leur pouvoir de suggestion en se basant sur le nombre de « followers », personnes abonnées à leur compte.

Problématique

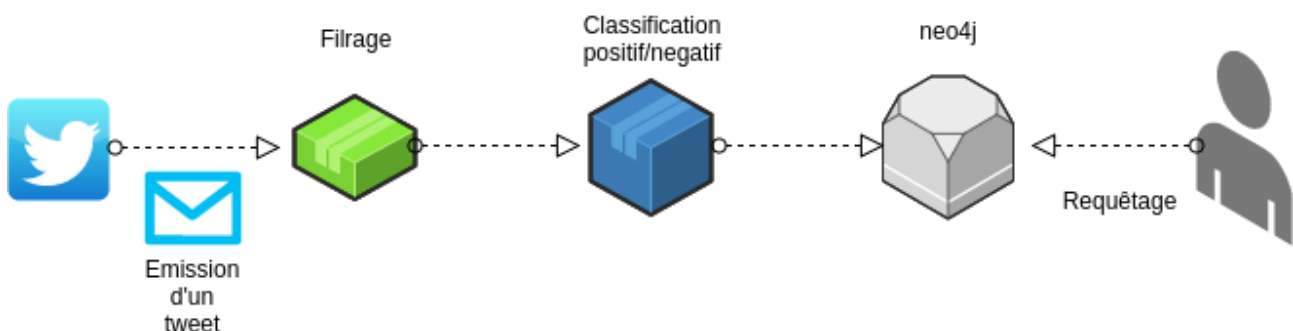
L'étude se propose d'étudier le réel pouvoir via la réaction des abonnés aux messages émis par les influenceurs. Pour ce faire, nous détecterons le type de réponse positive ou négative des abonnés. Puis, nous enrichirons notre base de ces informations en mettant en œuvre les connaissances acquises lors de cours tels que RCP208 ou RCP209. Nous ferons ensuite les relations entre influenceurs/abonnés via une analyse des liens à l'aide d'un graphe.

Nous nous limiterons principalement au réseau social Twitter. Cette plateforme présente l'avantage de fournir facilement les messages, ajoute de nombreuses méta-données (comme la position GPS) et le contenu fait moins de 280 caractères, obligeant à l'usage de termes forts.

Ce projet est fait dans le cadre du cours NFE204. Nous nous concentrons donc essentiellement sur la base de données NoSQL Neo4J. Il pourra être utilisé aussi dans le cadre d'autres cours comme RCP216 ou NFE211.

Présentation du projet

Architecture global



¹ https://en.wikipedia.org/wiki/Influencer_marketing

Etapes de l'analyse

Réception

Le serveur de réception des tweets sera développé en Python avec la librairie python-twitter. Nous récupérerons les données sur quelques jours afin d'avoir une base suffisante pour l'analyse.

Filtrage des tweets

Nous filtrerons les tweets selon la langue choisie (soit français, soit anglais).

- Sur la position GPS
- Sur la langue dans les méta-données
- Sur une analyse du message (détection de mots-clé, seuil, ...)

Transformation du tweet en une forme compréhensible

La partie transformation sera implémentée en Python avec la librairie NLTK. Nous mettrons en œuvre les fonctions suivantes :

- Stemmisation (racination)
- Lemmatization
- Suppression des tokens inutiles (stop words)

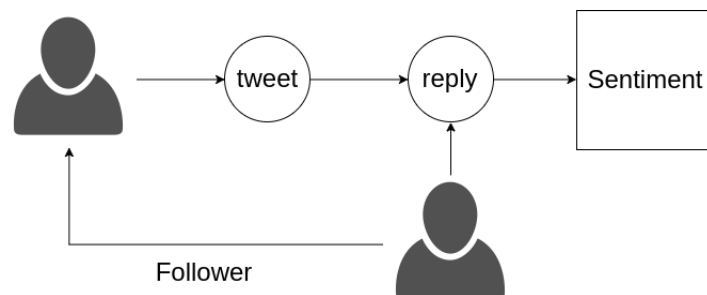
Enrichissement/Classification

La partie classification mettra en œuvre un algorithme de classification supervisé (Régression logistitique, SVM, arbres, réseaux de neurones, ...). On pourra réutiliser des analyses déjà effectuées sur le sujet telles que « Twitter Sentiment Analysis Training Corpus »².

- Utilisation d'une solution à base de word2vec pour construire un système de reconnaissance de l'aspect positif ou négatif du tweet
- Réutilisation d'une solution fiable

Base Neo4J

- Choix de la base
- Installation
- Description de la base, de son architecture, ...
- Construction du modèle. On pourra s'inspirer de modèles déjà existants³. Une première ébauche :



- Insertion des données

Recherche du véritable pouvoir des influenceurs

- Requêtage
 - Présentation du langage de requêtage sur une base de type “graphe”
- Analyse des résultats
- Limites et extensions

² <http://thinknook.com/twitter-sentiment-analysis-training-corpus-dataset-2012-09-22/>

³ <https://neo4j.com/blog/oscon-twitter-graph/>