IA

hackA tech

Shake science. Shape innovation.

#analyse

#connaissance

#langage

TALN
Comprendre et analyser le langage humain

Inria

CARACTÉRISTIQUES

Le traitement Automatique du Langage Naturel (TALN) ou Natural Language Processing (NLP) est un domaine pluridisciplinaire à la croisée de la linguistique et de l'intelligence artificielle (machine learning).

Ce domaine de recherche très dynamique permet aux machines (*ordinateurs*) de comprendre le langage naturel (*texte, parole*) produit par l'homme. Autrement dit, le TALN est en capacité de développer diverses applications pour analyser, traiter et reproduire le langage humain de manière automatique.

Ainsi pour réaliser de la traduction de texte, la correction de grammaire, des résumés automatiques de contenu, des synthèses vocales, ou des assistants virtuels (*chatbot*), les chercheurs ont développé des algorithmes spécifiques basés sur du Machine Learning et Deep Learning (*branches de l'Intelligence artificielle*).

Concrètement les applications de TALN sont en capacités de réaliser des traitements :

- **Sémantiques** : génération de contenu, correction, traduction,...
- **Syntaxique :** regroupement de mots séparation des phrases d'un texte,...
- **De Signal (parole & graphie) :** détection des langues, analyse écriture manuscrite,...

QUELS AVANTAGES ?

- Analyser et comprendre le langage naturel (*texte, parole*) automatiquement grâce aux machines
- Traiter de façon exhaustive des contenus extrêmement volumineux par du Machine Learning
- Améliorer de façon continue les différentes analyses

© Sergey Nivens - Fotolia

TRAITEMENT DES DONNÉES

Des contenus annotés comme des documents écrits (*textes, phrases, sous-mots*) ou sonores.

Des libraires de données sont actuellement disponibles comme Spacy (*modèles pré-entrainés*).

USE CASES

- **Marketing :** analyse automatique de retours clients, modération automatique de forums
- **Santé :** analyse automatique de rapports de R&D pour détection de cas particuliers
- **Juridique & Commerce :** analyse automatique d'appel d'offres
- **Presse :** génération automatique de résumés

FICHE IDENTITÉ

- Équipe projet : Linkmedia

FONCTIONNALITÉS GÉNÉRIQUES

- **Régression logisitique :** Pour l'analyse des sentiments, les analogies complètes, les traductions de textes.
- **Locality sensitive hashing (LSH) :** pour l'approximation des plus proches voisins.
- **Programmation dynamique, Modèle de Markov caché :** pour la correction automatique de mots mal orthographiés ou de phrases incomplètes, l'étiquetage morpho-syntaxique (*Part Of Speech (POS) Tagging*).
- **Réseaux de neurones récurrents, LSTM, GRU et réseaux de neurones Siamois dans TensorFlow et Trax :** pour l'analyse avancée des sentiments, la génération de textes, la reconnaissance des entités, l'identification de questions dupliquées.
- **Encodeur décodeur, modèles T5, BERT :** pour la traduction avancée de textes complets, le résumé de textes, les questions réponses, la construction de chatbots.

RÉGRESSION LOGISITIQUE

La régression logistique est une approche statistique qui peut être employée pour évaluer et caractériser les relations entre une variable réponse de type binaire (*par exemple : succès / échec, malade / non malade*), et une, ou plusieurs, variables explicatives, qui peuvent être de type catégoriel (*par exemple : le sexe*), ou numérique continu (*par exemple : l'âge*).

Tout comme la régression de Poisson, la régression logistique appartient aux modèles linéaires généralisés. Pour rappel, il s'agit de modèles de régression qui sont des extensions du modèle linéaire, et qui reposent sur trois éléments :

1. un prédicteur linéaire
2. une fonction de lien
3. une structure des erreurs

PROGRAMMATION DYNAMIQUE

la programmation dynamique permet de résoudre tout problème d'optimisation dont la fonction objectif se décrit comme la somme de fonctions monotones non-décroissantes des ressources. Concrètement, cela signifie que l'on va pouvoir déduire la solution optimale d'un problème à partir des solutions optimales de tous les sous problèmes.

Les problèmes d'optimisation dynamique ont des applications importantes aussi bien dans l'industrie qu'en gestion. Il s'agit de minimiser le coût d'une trajectoire dans un espace d'états. On dispose d'une loi d'évolution, qui détermine l'état suivant à partir de l'état courant et d'une << commande >> ; les trajectoires sont construites à partir d'un état initial et d'une suite de commandes, suivant cette loi d'évolution ; on se donne également une fonction d'objectif, définie sur les trajectoires, qu'il s'agit de minimiser.

CONNAISSANCES MINIMUM REQUISES

- **Algorithmes de Machine Learning et Deep Learning**
- **Exploration de données**
- **Méthodes statistiques**.

Référent : Vincent Claveau

* Linkmedia est une équipe-projet commune à Inria, au CNRS, INSA Rennes et Université de Rennes 1.

INSTITUT CARNOT Inria

UNION EUROPÉENNE

l'Europe s'engage en France

Inria