

Projet de synthèse

Réseaux de neurones artificiels : application à la classification des sentiments des discussions en ligne

Le problème de la fouille de texte a suscité de plus en plus d'attention dans les dernières années en raison de grandes quantités de textes disponibles sur les forums web, les réseaux sociaux, et autres types d'applications.

La caractéristique la plus importante des données textuelles est leur grande dimensionnalité et leur caractère éparé et hétérogène. Par exemple, un corpus de données peut être établi à partir d'un lexicon de 100.000 mots ; un document peut aussi ne contenir qu'une centaine de mots.

Les données auxquelles on s'intéresse sont des matrices croisant documents et mots. Ces documents peuvent être issus de forums de discussion où chaque document peut représenter une discussion sur un thème quelconque. Ce sont des données de grande taille, très éparées, où le nombre d'attributs (mots) dépasse largement le nombre d'individus (documents). Les méthodes de classification simple en général ne donnent pas de bons résultats sur ce type de données.

Le problème devient plus difficile quand d'autres sources de données doivent être prises en compte, notamment des images qui représentent des visages.

Dans le cadre de ce projet, le flux de données de type texte sera modélisé par des méthodes de clustering à base de réseaux de neurones.

Le travail sera fait en langage Python avec les bibliothèques numpy et scikit-learn.



Figure 1. Détection des opinions à partir d'images et du texte

La figure 1 représente un exemple de détection des opinions/émotions à partir de données textuelles et visuels qui concerne la même population.

Résultats attendus:

Développement d'un système informatique en utilisant une méthode d'apprentissage semi-supervisé capable de traiter ce types de données (images et texte) et retourne une classification des discussions/échanges à en utilisant ces données hétérogènes.

Compétences souhaitées :

- Traitement de données
- Développement Python et /ou R

Contact

Nistor GROZAVU : nistor.grozavu@cyu.fr

Dimitris KOTZINOS : Dimitrios.Kotzinos@cyu.fr