

L'INTELLIGENCE ARTIFICIELLE DES TEXTES

Des algorithmes à l'interprétation

Sous la direction de Damon MAYAFFRE et Laurent VANNI



PARIS
HONORÉ CHAMPION ÉDITEUR
2021

www.honorechampion.com

TABLE DES MATIÈRES

| | |
|--------------------------|---|
| <i>IN MEMORIAM</i> | 7 |
|--------------------------|---|

| | |
|---------------------------|---|
| PRÉSENTATION | 9 |
|---------------------------|---|

Laurent VANNI et Frédéric PRECIOSO

| | |
|--|----|
| DEEP LEARNING ET DESCRIPTION DES TEXTES ARCHITECTURE MÉTHODOLOGIQUE | 15 |
|--|----|

| | |
|-------------------|----|
| Introduction..... | 15 |
|-------------------|----|

| | |
|--------------------|----|
| 1. Prérequis | 18 |
|--------------------|----|

| | |
|-------------------------------------|----|
| 1.1. Le traitement des données..... | 18 |
|-------------------------------------|----|

| | |
|-------------------------------------|----|
| 1.2. L'architecture du modèle | 21 |
|-------------------------------------|----|

| | |
|--|----|
| 1.3. Les (hyper)paramètres du réseau | 25 |
|--|----|

| | |
|---|----|
| 2. <i>Embedding</i> : représentation des mots | 32 |
|---|----|

| | |
|--|----|
| 3. Convolution : abstraction des données | 39 |
|--|----|

| | |
|---|----|
| 3.1. <i>Text Deconvolution Saliency</i> (TDS) | 41 |
|---|----|

| | |
|------------------------------|----|
| 3.2. Pondération du TDS..... | 51 |
|------------------------------|----|

| | |
|---|----|
| 4. Analyse multi-channels : formes, lemmes, étiquettes morpho- syntaxiques | 55 |
|---|----|

| | |
|-------------------|----|
| 4.1. Modèle | 55 |
|-------------------|----|

| | |
|----------------------|----|
| 4.2. Exemple 1 | 58 |
|----------------------|----|

| | |
|----------------------|----|
| 4.3. Exemple 2 | 59 |
|----------------------|----|

| | |
|----------------------|----|
| 4.4. Remarques | 61 |
|----------------------|----|

| | |
|--|----|
| 5. Les entrées et sorties du réseau..... | 62 |
|--|----|

| | |
|--|----|
| 5.1. Détection des passages-clés | 62 |
|--|----|

| | |
|--------------------------------|----|
| 5.2. Filtrage des données..... | 66 |
|--------------------------------|----|

| | |
|------------------|----|
| Conclusion | 70 |
|------------------|----|

ÉTIENNE BRUNET, LUDOVIC LEBART ET LAURENT VANNI

| | |
|--|-----------|
| LITTÉRATURE ET INTELLIGENCE ARTIFICIELLE | 73 |
| Introduction..... | 73 |
| 1. Le roman au XX ^e siècle. Reconnaître que deux textes sont d'un même auteur..... | 75 |
| 1.1. Une expérience en double aveugle | 75 |
| 1.2. Les mesures classiques de la distance intertextuelle..... | 77 |
| 1.3. Méthodes fondées sur le TLE | 80 |
| 1.4. La solution du <i>deep learning</i> | 89 |
| 2. Le théâtre classique..... | 94 |
| 2.1. Les basses fréquences | 96 |
| 2.2. Les hautes fréquences | 101 |
| 2.3. Neutralisation du genre | 103 |
| 2.4. Le <i>Deep learning</i> | 104 |
| 3. Essai d'explicitation et d'intégration | 107 |
| 3.1. Imitation de la procédure du <i>deep learning</i> : les triplets.... | 108 |
| 3.2. Réduction et intégration des deux approches, <i>procédurale et algorithmique</i> | 114 |
| 3.3. La déconvolution | 119 |
| 3.4. Analyse non-supervisée des « pages » (fragments de 50 lignes consécutives) | 125 |
| Conclusion (Éléments pour une...)..... | 128 |

MAGALI GUARESI ET DAMON MAYAFFRE

| | |
|---|------------|
| INTELLIGENCE ARTIFICIELLE ET DISCOURS POLITIQUE..... | 131 |
| Introduction..... | 131 |
| 1. <i>Deep learning</i> et textes politiques profonds. Cadrage | 135 |
| 1.1. Convolution, co-texte, cooccurrences | 136 |
| 1.2. Réseaux, couches cachées, complexité..... | 137 |
| 1.3. Du sens en contexte | 138 |
| 2. Vers une description des discours gauche / droite. Les professions de foi des députés sous la V ^e République | 141 |
| 2.1. Gauche et droite décrites par <i>deep learning</i> (1958-2017).. | 143 |
| 2.2. Situer les professions de foi macronistes sur l'échiquier politique | 156 |

| | |
|--|-----|
| 3. Vers une objectivation de l'intertexte. Les emprunts de Macron à ses prédécesseurs à l'élysée | 162 |
| 3.1. Proposition méthodologique : un usage original du <i>deep learning</i> | 165 |
| 3.2. Résultats : intertexte pluriel et discours patchwork | 167 |
| Conclusion | 180 |

DOMINIQUE LONGRÉE

| | |
|--|-----|
| MOTIFS TEXTUELS ET DEEP LEARNING : VERS UNE DÉTECTION AUTOMATIQUE DE NOUVEAUX OBSERVABLES LINGUISTIQUES ? | 183 |
| Introduction..... | 183 |
| 1. Motifs textuels et unités phraséologiques. Critères d'identification | 186 |
| 2. Motifs textuels et détections semi-supervisées | 189 |
| 2.1. Motifs et Hyperbase Web Edition | 189 |
| 2.2. SDMC-Sequential Data Mining under Constraints | 190 |
| 3. Motifs textuels et hyperdeep..... | 191 |
| 4. Hyperdeep : un outil heuristique pour le latin ?..... | 200 |

FRANÇOIS RASTIER

| | |
|---|-----|
| DATA VS CORPORA | 203 |
| 1. Un bref retour sur le mythe de l'IA | 203 |
| 2. Rien ne nous est donné..... | 208 |
| 3. Traitements et maltraitements | 219 |
| 4. Limites de l'émergentisme..... | 222 |
| 5. Pourquoi les corpus sont indispensables..... | 230 |
| 6. Le problème des passages et l'expérimentation connexionniste. | 238 |
| 7. La <i>Data Science</i> contre les sciences de la culture..... | 244 |
| BIBLIOGRAPHIE | 247 |
| INDEX DES PERSONNES | 257 |
| TABLE DES MATIÈRES | 263 |