Hyperdeep : deep learning descriptif pour l’analyse de données textuelles

Laurent Vanni[1], Marco Corneli[2], Dominique Longrée[3], Damon Mayaffre[4], Frederic Precioso[5]

[1]Université Côte d’Azur, CNRS, BCL (UMR 7320), France – laurent.vanni@univ-cotedazur.fr

[2]Université Côte d’Azur, Center of Modeling, Simulmation & Interaction, France – marco.corneli@univ-cotedazur.fr

[3]Université de Liège, L.A.S.L.A, Belgique – dominique.longree@uliege.be.

[4]Université Côte d’Azur, CNRS, BCL (UMR 7320), France – damon.mayaffre@univ-cotedazur.fr

[5]Université Côte d’Azur, CNRS, I3S (UMR 7271), France – Frederic.Precioso@univ-cotedazur.fr.

Abstract 1 (in English)

Since few years, some tools that are helping us to interpret results of deep learning have appeared (LIME, LSTMVIS, TDS). In this paper, we propose to go further by searching hidden information encoded in intermediate layers of deep learning thanks to a new tool. Hyperdeep allows, on the one hand, to predict the belonging of a text and to appreciate its borrowings from different styles or authors and, on the other hand, it allows to analyze, by deconvolution, the textual saliency in order to bring up the linguistic markers learned by the network. This new type of linguistic objects is gathered and highlighted in a graphical tool combining visualizations and hypertext. This tool is fully integrated in the Hyperbase Web platform, which offers the adequate environment and a natural starting point for any study mixing deep learning and text mining.

Keywords: deep learning, text-mining, classification, prediction, description

Abstract 2 (in French)

Depuis peu, les outils d'aide à l'interprétation des résultats du deep learning font leur apparition (LIME, LSTMVIS, TDS). Dans cette communication nous proposons d'aller plus loin en allant chercher l'information cachée au plus profond des couches intermédiaires du deep learning grâce à un nouvel outil. Hyperdeep permet d'une part de prédire l’appartenance d’un texte et d’en apprécier les emprunts à différents styles ou auteurs et d’autre part, par déconvolution, d'analyser les saillances du texte afin d’en faire remonter les marqueurs linguistiques appris par le réseau. Cette information d’un genre nouveau est rassemblée et mise en valeur dans un nouvel outil mêlant visualisations graphiques et texte dynamique. Son utilisation est accompagnée d’une intégration complète dans la plateforme Hyperbase Web qui propose l’environnement adéquate et un point de départ naturel pour toute étude mêlant deep learning et statistiques du texte.

Mots clés : deep learning, text-mining, classification, prédiction, description

1. Introduction

Les méthodes d'apprentissage utilisant les réseaux de neurones artificiels sont depuis longtemps considérées comme des généralisations des méthodes statistiques plus traditionnelles. Une analyse factorielle des correspondances peut par exemple être considérée comme un réseau de neurones particulier (linéaire) à une seule couche cachée (Lebart, 1997). Cette généralisation associée à la puissance de calcul des machines modernes conduit aujourd'hui le deep learning à atteindre des performances inédites en tâches de prédictions (classifications supervisées). Malheureusement ces modèles introduisent une complexité algébrique telle que la description des données traitées devient presque impossible. L'ADT se retrouve confrontée au choix de la méthode, avec une approche descriptive des textes dominée par la statistique transparente en termes de calculs, et une approche prédictive maitrisée par le deep learning mais fonctionnant en boîte noire.

Heureusement des méthodes existent aujourd'hui pour visualiser et tenter d'interpréter la mécanique interne du deep learning (Montavon et al. 2018). Certaines méthodes s'appliquent à travailler sur les entrées et sorties des réseaux de neurones pour en déduire l'importance de chacun des éléments traités par l'IA. C'est le cas par exemple de LIME (Ribeiro et al., 2016) qui permet d'interpréter n'importe quel classifieur, en donnant un indice de participation à chaque mot pour la prise de décision du modèle. D'autres méthodes analysent directement les couches cachées pour observer les taux d'activation de chaque neurone et donner un score à chaque token ayant participé au calcul final d'attribution de la classe. Nous pouvons citer ici LSTMVis (Strobelt et al., 2018) pour les réseaux récurrents ou encore le modèle TDS[1] (Vanni et al., 2018a) pour les réseaux convolutionnels.

Entre prédiction et description, l'outil que nous proposons dans cette contribution, Hyperdeep, reprend le modèle TDS et propose une interface visuelle nouvelle afin d'en exploiter les moindres détails et d'interpréter au mieux les indices descriptifs remontés par le réseau[2]. Pour y parvenir, nous présenterons dans un premier temps une variante du TDS originel qui permet de répondre aux besoins précis de l'analyse des textes. Nous verrons ensuite comment l'interface graphique, intégrée à la plateforme Hyperbase Web[3], met à la portée des chercheurs de nouveaux observables linguistiques issus de l'apprentissage du réseau de neurones. Enfin nous étudierons deux cas concrets d'utilisation d'Hyperdeep qui illustrent la plus-value de cet outil dans la continuité des méthodes d'analyses traditionnelles d'ADT (Lebart, Pincemin, Poudat, 2019) desquelles Hyperdeep s'inspire directement.

2. Modèle

Les réseaux de neurones artificiels offrent de nombreux types de modèles différents. Parmi eux, on peut distinguer deux catégories. Les réseaux récurrents d'un côté et les réseaux convolutionnels de l'autre. Chacun de ces réseaux propose des spécificités qui le rendent performant pour un type de tâche donnée. Les réseaux récurrents ont été implémentés pour être sensibles aux données de type séquentiel ou temporel. Chaque neurone ou couche de neurones conserve une partie de la mémoire d'une séquence et la communique à la suivante moyennant une fonction d'activation (vraisemblablement non linéaire). Cette particularité les

1 TDS : Text Deconvolution Saliency

2 Ce travail a bénéficié d'une aide du gouvernement français, gérée par l'Agence Nationale de la Recherche au titre du projet Investissements d'Avenir UCAJEDI portant la référence n° ANR-15-IDEX-01

3 Plateforme d'analyses de données textuelles en ligne : hyperbase.unice.fr

rend sensibles à la modélisation de la séquence. Dans le cas du texte, les réseaux récurrents sont performants pour modéliser une langue, apprendre des règles de construction d'une phrase ou d'une syntaxe et prédire en sortie le prochain caractère, mot ou token d'une séquence textuelle. En ADT classique, notre approche du texte est un peu différente. Nous travaillons généralement en contraste : les logiciels tel qu'Hyperbase s'appuient sur des métadonnées[4] pour partionner le corpus et de décrire les textes en comparaison[5]. Cette logique classificatoire est le point de départ qui nous pousse ici à nous tourner vers le deuxième type de modèle, plus performant pour la classification de textes, les réseaux convolutionnels.

Inspiré du cortex visuel, ce type de réseau considère le texte comme une donnée spatiale statique (une image) et interprète les saillances textuelles dans leur contexte (les contours de l'image). Le texte est balayé par une multitude de filtres qui vont se spécialiser au fil de l'apprentissage (les filtres sont des paramètres du réseau modifiés par back propagation) pour obtenir un modèle abstrait du texte qui ne conserve que les saillances textuelles pertinentes pour la tâche d'apprentissage donnée. Cette méthode bio-inspirée engendre des classifieurs supervisés redoutables qui dépassent dans certains cas les méthodes statistiques classiques (Brunet, et al., 2020 ; Brunet et al., 2019).

La réelle plus-value pour l'ADT de ce type de méthode est de nous offrir une description des données d'un genre nouveau. Nous avons vu avec (Vanni et al., 2018) que nous pouvions lire et interpréter les filtres de la convolution en utilisant une couche supplémentaire de déconvolution qui permet le calcul du TDS. Cependant beaucoup de points restent à éclaircir quant à la participation de chaque marqueur linguistique identifié par le réseau pour la classification. Pour répondre à cette problématique, nous proposons d'utiliser une variante du calcul du TDS, le TDS pondéré (Vanni et al., 2020 – soumis).

Le TDS tel qu'il a été proposé en 2018 correspond au score d'activation de chaque mot en sortie de convolution. Ce score est en fait calculé comme une somme d'activations de neurones associés à chacun des mots. Le problème de cette méthode est qu'elle ne prend pas en compte le reste du réseau profond. Après la convolution, il y a au moins une couche de neurones type MLP (Multi Layer Perceptron) qui pondère ce calcul avant d'activer la dernière couche du réseau (la couche de sortie) voir Figure 1.

Cette pondération prise en compte dans le calcul permet d'ajuster, voir d'inverser, la valeur de chacun des marqueurs linguistiques appris par le réseau. Cet ajustement a un impact majeur sur l'interprétabilité des résultats. Un TDS élevé n'est plus seulement un marqueur important pour la classification en général. Il peut être maintenant directement associé à une classe en particulier. Avec cette méthode chaque token a un poids différent selon la classe observée. Ce qui rend l'utilisation du TDS beaucoup plus pertinente pour le linguiste qui peut ainsi

4 Le terme métadonnée en ADT est comparable à celui de classe utilisé dans une classification et en particulier en deep learning.

5 La plupart des calculs statistiques utilisés en ADT demandent au minimum deux textes comme par exemple le calcul des spécificités. Ce minimum peut même atteindre quatre textes dans le cas d'analyses plus complexes comme la classification arborée ou encore l'Analyse Factorielle des Correspondances (AFC).

concentrer son interprétation sur les marqueurs linguistiques du texte qui ont réellement permis de l'attribuer.

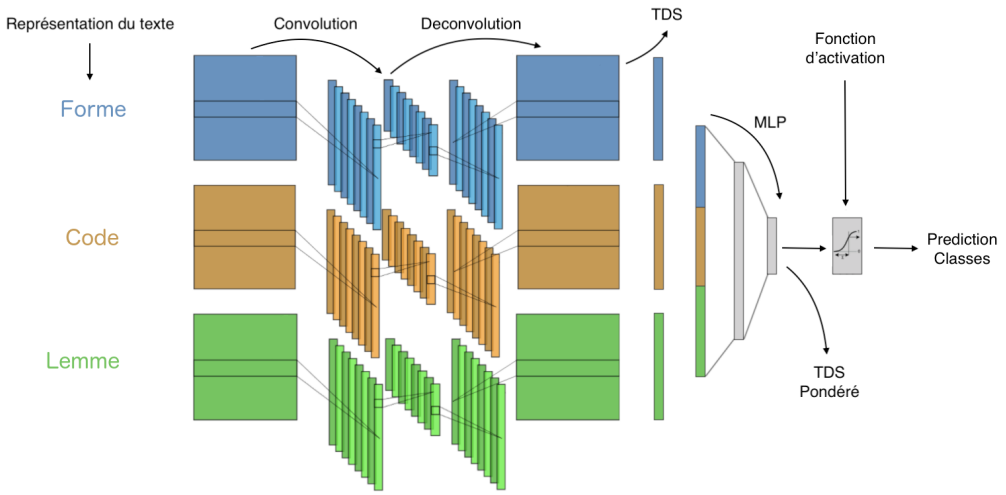

Figure 1 : Modèle deep learning de convolution/déconvolution implémenté par Hyperdeep.

La lemmatisation est un autre aspect qui vient enrichir le TDS tel que nous le présentons dans Hyperdeep. Les modèles convolutionnels acceptent très bien l'utilisation de plusieurs *channels* (plusieurs représentations des données à la fois). Ces *channels* sont un moyen confortable d'introduire un étiquetage morphosyntaxique en entrée du réseau. Chaque étiquette est alors considérée comme un token qui peut être aussi observé par le TDS. Cette solution apporte une couche supplémentaire d'information linguistique qui alimente l'interprétation des sorties du réseau par l'humain.

La déconvolution n'est pas la seule responsable de la plus-value heuristique du deep learning en ADT. La dernière couche du réseau livre, elle aussi, une quantité d'information non négligeable. Elle est communément utilisée pour extraire une simple probabilité, celle d'identifier l'une ou l'autre des classes. Pour y parvenir, cette dernière couche d'activation contraint le résultat numérique dans un ensemble compris entre 0 et 1. Or chaque passage du texte étudié en entrée du réseau, même s'il correspond à la même classe, provoque un taux d'activation potentiellement différent sur la couche de sortie. Malheureusement cette information est perdue avec l'utilisation de fonctions d'activations telles que le *softmax* qui écrase toutes les valeurs en sortie. L'idée est donc ici d'observer l'activation du réseau juste avant la sortie et ainsi de capter toutes les variations (Figure 1). Avec cette méthode il est possible de hiérarchiser l'information et ainsi de connaitre les extraits qui ont le plus fortement activé le réseau. Cette méthode est d'ailleurs suffisamment robuste pour pouvoir accepter de ressoumettre les données d'entrainement pour en observer les passages-clés (Vanni et al., 2018b). Une pratique jusqu'ici interdite par la communauté du deep learning[6] mais qui prend ici tout son sens grâce au caractère descriptif de cette méthode et du TDS.

6 En deep learing les données d'entrainement sont strictement réservées à l'apprentissage. Seules de nouvelles données (validation) peuvent être utilisé en prédiction notamment pour calculer la précision (*accuracy*) ou la généralisation (*loss*) du modèle.

Hyperdeep propose donc plusieurs cas d'utilisation possibles, il peut être soit simplement prédictif, soit complètement descriptif avec l'analyse de chacune des classes (ou métadonnée) du corpus et l'identification des passages-clés. C'est aussi un moyen de capter l'intertexte lorsqu'on retire de l'apprentissage une partie (une classe) du corpus et qu'on la ressoumet ensuite à la prédiction pour observer l'emprunt de cette partie aux autres (Mayaffre et al., 2020)

Le deep learning vient donc compléter l'offre descriptive des modèles statistiques sans pour autant être concurrent. Les modèles convolutionnels permettent d'observer des objets linguistiques d'un genre nouveau et de ce fait un aller-retour entre statistique et deep learning est souvent nécessaire pour confirmer une intuition linguistique suggérée par la machine. Nous avons donc choisi d'intégrer Hyperdeep à la plateforme d'analyse statistique de donnée textuelle Hyperbase Web. Ce logiciel en ligne représente ainsi une sorte de bac à sable pour étudier le comportement du deep learning et comparer les sorties machines à celles de nos outils statistiques historiques. Nous allons voir maintenant quelles solutions techniques et graphiques nous proposons pour rendre l'utilisation du deep learning accessible et interprétable pour les chercheurs en ADT.

3. Interface

3.1. Préparation des données

3.1.1. Le corpus et ses méta-données

Hyperbase Web est une plateforme en ligne d'ADT qui propose une suite d'outils statistiques qui fonctionnent pour la plupart sur la base d'analyses dites contrastives. Il est donc nécessaire de préparer en amont ses données avec un partitionnement pertinent qui représente une hypothèse de travail. Cette approche d'étiquetage du corpus est essentielle aux analyses statistiques, mais aussi le point de départ de tout apprentissage supervisé et a fortiori en deep learning. La préparation des données est donc la même qu'en ADT, il n'y a pas de surcoût de préparation des données pour utiliser Hyperdeep lorsque l'on utilise déjà un logiciel tel qu'Hyperbase, Lexico, TXM ou encore Iramuteq. De plus Hyperbase Web propose déjà un mécanisme qui permet de moduler, fusionner, rassembler et créer un partitionnement souple à partir de métadonnées. Ce partitionnement est la clé d'un apprentissage réussi.

La taille joue bien évidemment un rôle important sur la précision du modèle final. Le deep learning est gourmand en termes de nombre de mots, mais on peut distinguer deux cas de figure qui impliquent des tailles minimums de corpus différentes. Pour un partitionnement de type documentaire (factuel), par exemple sur des noms d'auteurs ou sur des dates, le deep learning s'en sort en général bien. Environ 100 000 mots par partie peuvent alors suffire à la machine pour créer une représentation abstraite des données et obtenir un modèle avec un bon taux de précision et de généralisation. En revanche pour un partitionnement de type hypothétique (à confirmer par l'analyse) comme par exemple définir la gauche et la droite en France sous la 5e république, plusieurs millions de mots sont nécessaires pour obtenir un apprentissage satisfaisant.

Il faut être aussi sensible aux biais dans le corpus. Mais heureusement ce sont les mêmes biais qui peuvent affecter la statistique et être détectés par un simple calcul de spécificité ou encore par une AFC (par exemple certains noms propres, certaines dates et autres spécificités fortes). L'intégration d'Hyperdeep dans Hyperbase permet donc de contrôler finement ces questions et faire toutes les vérifications nécessaires avant un apprentissage. Une fois le corpus nettoyé il ne reste plus qu'à lancer un entrainement. Il est alors souvent nécessaire de calibrer certains paramètres du deep learning pour s'adapter aux différents cas de figure.

3.1.2. L'entrainement deep learning

Hyperdeep est accessible depuis le menu déroulant d'Hyperbase Web et propose, comme point d'entrée, de lancer un apprentissage sur le partitionnement courant. L'architecture des réseaux de neurones profonds repose sur un ensemble de paramètres qui modifient la complexité ou la profondeur des réseaux, les hyperparamètres. Ces paramètres peuvent potentiellement affecter la qualité de l'apprentissage. La plupart (*bach size*, *learning rate*, *epoch*, ...) sont accessibles via les options de l'interface graphique, mais sont en général réservés aux utilisateurs avertis du deep learning.

Cependant qu'il soit averti ou non, il n'existe pas de règle ou de méthode pour prédire le paramétrage optimal de chaque modèle. Il faut tester, corriger et retester encore jusqu'à atteindre un état stable convenable. Cette façon de construire une IA contribue à véhiculer l'image de boite noire du deep learning où les hyperparamètres sont une sorte de recette de cuisine qui est propre à chaque apprentissage. Néanmoins, certaines bonnes pratiques existent et peuvent être appliquées pour modifier raisonnablement ces paramètres. Beaucoup d'articles existent à ce sujet (Aghaebrahimian et al 2019), nous ne détaillerons donc pas ici comment paramétrer un réseau de neurones, mais il faut noter que le taux de précision d'un modèle par défaut peut significativement évoluer en modifiant les options d'Hyperdeep avant de lancer un l'apprentissage. Une fois l'apprentissage terminé le score de précision du modèle est affiché (*accuracy* et *loss*) et l'ensemble des fonctionnalités d'Hyperdeep est activé.

3.2. Affichages des résultats

3.2.1. Prédiction

La première fonctionnalité disponible après un entrainement est la classification de texte, une prédiction de l'appartenance d'un texte nouveau (inconnu du système lors de l'apprentissage) à l'une des classes. Hyperdeep fonctionne par défaut avec un simple copier-coller, mais il est possible de charger un fichier texte volumineux qui sera alors analysé dans son ensemble. Comme les réseaux convolutionnels fonctionnent sur des données de taille fixe (voir section 2) le texte présenté sera découpé en segments de taille fixe, la longueur de ces segments étant un des nombreux hyperparamètres à choisir avant l'apprentissage. Chaque segment est ainsi attribué indépendamment des autres à une des classes. Et la somme de ces attributions donne un score global réparti entre toutes les classes. Cette répartition donne une idée du profil général du texte. Pour accompagner l'interprétation de cette classification, une visualisation graphique est proposée avec un diagramme de type radar qui permet d'avoir un aperçu rapide de l'orientation d'un texte (Figure 2) .

Cette tâche de prédiction est naturelle pour la communauté du deep learning. Les différents cas d'utilisations en SHS reste encore à découvrir et sont maintenant entre les mains des chercheurs. Par exemple il est possible, comme mentionné en introduction, d'analyser un nouveau texte pour l'attribuer aux différentes classes (attribution d'auteur par exemple), ou

encore de retirer une partie de corpus (un auteur) pour le soumettre à la prédiction et apprécier l'intertexte (Mayaffre 2020 - soumis). Et il y a certains cas qui ne sont possibles que par le biais des passages-clés et en utilisant la plus-value descriptive du deep learning (ici le TDS). À savoir, redonner tout le corpus d'entrainement à la tâche de prédiction pour avoir une description fine des caractéristiques linguistiques fortes de chaque classe identifiées par les couches cachées du réseau. Nous allons voir maintenant comment exploiter cette description dans Hyperdeep

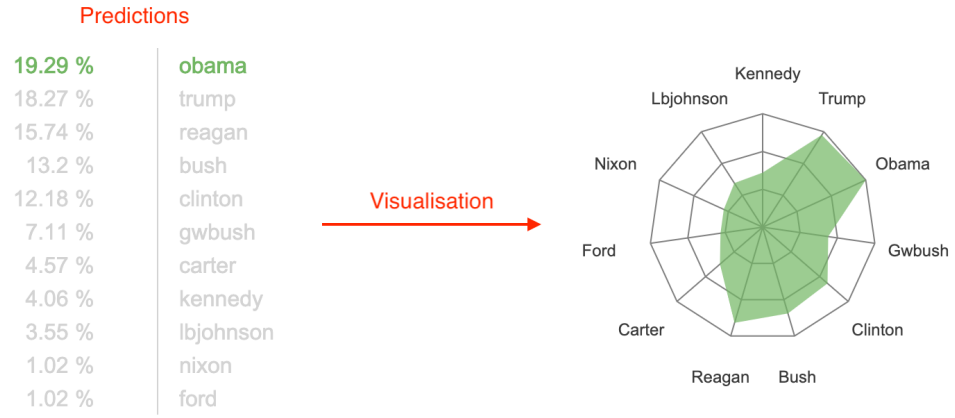

Figure 2 : Visualisation d'une prédiction dans Hyperdeep pour un texte donné découpé en segments de taille fixe (L'ensemble des segments est égal à 100%)

3.2.1. Description

Le calcul du TDS (pondéré) attribue un score à chaque mot (chaque token) pour chaque classe. Sa participation à la prise de décision finale peut être ainsi évaluée et comparées dans le contexte du partitionnement du corpus. Et avec l'utilisation de la lemmatisation proposée par Hyperbase et réutilisée par Hyperdeep chaque mot correspond en fait à trois calculs de TDS, un pour chaque représentation du mot : la forme graphique, le code grammatical et le lemme. Ces représentations sont entièrement contextualisées par la convolution qui applique une fenêtre glissante sur le texte et qui les regroupes ensemble dans les dernières couches du réseau. Le TDS du mot est donc dépendant de son contexte et un passage devient clé pour le réseau au regard de l'ensemble des informations (forme/code/lemme) présent dans le passage.

Pour chaque token, le TDS peut être soit positif soit négatif selon la classe observée en sortie et selon si le token a servi ou au contraire desservi cette classe. Pour capturer l'ensemble de ces informations et proposer un parcours interprétatif intuitif au chercheur, nous proposons une visualisation dynamique du texte (voir Figure 3). Trois couleurs ont été choisies, bleu, orange, vert pour distinguer respectivement la forme, le code et le lemme. Par défaut un seuil de significativité est choisi pour n'afficher que les TDS les plus élevés (à la manière du seuil de spécificité de la loi normale par exemple), mais ce seuil peut être ajusté par l'utilisateur au moyen d'un curseur qu'il peut déplacer pour chaque représentation du mot. Le côté réellement dynamique de cet affichage vient du fait que l'on peut survoler chacune des classes pour voir s'afficher les TDS les plus élevées de la classe sélectionnée. Même si un passage est attribué à une classe, chaque mot a un poids différent selon la classe observée. Cette méthode permet de repérer notamment les caractéristiques partagées (ou pas) par plusieurs classes en même temps et de comprendre finalement pourquoi le passage a été attribué ou pas à une classe.

Il reste la notion de passages-clés qui tient une place importante dans l'interface d'Hyperdeep. Le texte peut naturellement être parcouru de page en page du début à la fin, mais il est possible d'aller aussi directement aux passages-clés de chaque classe (le segment de taille fixe le plus fortement attribué à la classe). Cette visualisation permet de se focaliser sur les parties du texte où le réseau de neurones a le plus de certitudes (une probabilité élevée pour la classe repérée). La hiérarchisation de l'information permise par le calcul des passages-clés donne un aperçu des saillances textuelles et autres marqueurs linguistiques les plus forts dans le texte. C'est avec cette méthode qu'il est possible de reparcourir tout un corpus d'entrainement pour obtenir non plus une prédiction mais bien une description de chaque classe et découvrir les segments de texte les plus caractéristiques et en apprécier les motifs linguistiques repérés par le réseau.

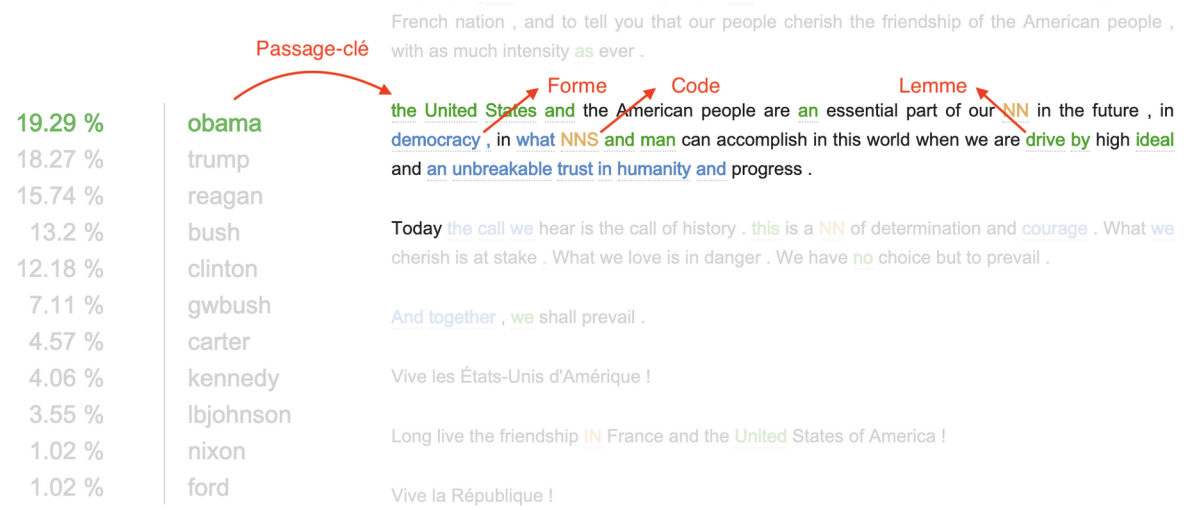

Figure 3 : Affichage des segments les plus caractéristiques et description des passages-clés dans Hyperdeep.

Hyperbase étant polyglotte, nous allons maintenant voir deux cas concrets d'utilisation d'Hyperdeep sur deux langues différentes, l'anglais et le latin qui présentent toutes deux certaines caractéristiques différentes que le réseau deep learning est capable d'apprécier.

4. Exemples

4.1. Trump, Obama… Macron : les résonances du discours

En s'adressant, en anglais, au Congrès des Etats-Unis le 25 avril 2018, Emmanuel Macron réalise un exercice difficile. Le protocole diplomatique, les contraintes politiques, la bien-saillance exigée de l'invité contraignent évidemment le discours. Et la presse américaine est suspendue à la tonalité d'un discours qui sera jugé favorable ou défavorable aux Démocrates ou aux Républicains.

Les couches cachées du deep learning et le traitement statistique permettent de montrer comment le discours du président français s'applique à reprendre majoritairement le discours démocrate d'Obama pour faire passer un message universaliste et écologique, tout en reprenant la forme du discours actuel des Républicains et de Trump.

Dans un vaste corpus d'apprentissage composé des discours des présidents américains depuis Kennedy, l'algorithme indique ainsi que le discours de Macron emprunte d'abord à Obama (19.29%). Par exemple, un des passages clés attribué à Obama résonne de mots que le

président démocrate aimait en effet utiliser et que le wTDS surligne (figure 4). La thématique écologique renvoie explicitement à la présidence Obama (et s'oppose au refus de Trump de ratifier les accords de Paris). Statistiquement, par exemple, les mots « climate », « planet » ou « transition » sont, de fait, caractéristiques d'Obama dans l'histoire américaine (figure 5)

Pour autant, Hyperdeep détecte aussi l'empreinte de Trump (18.27%) sur le discours que prononce Macron, notamment dans la forme hyperbolique du discours (adverbes et adjectifs forts) et dans une rhétorique relâchée, de type oral, souvent centrée sur la personne.

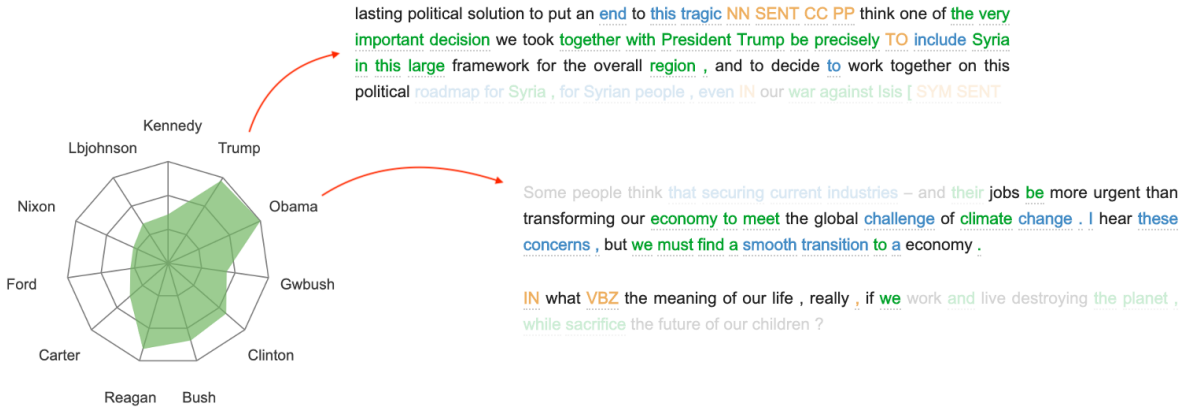

Figure 4 : Passages-clés attribués à Trump et Obama dans le discours de Macron.

Le TDS peut être confirmé par la distribution statistique des observables du passage : par exemple, comme son hôte américain aime le faire, Macron utilise ici des formules superlatives comme « very important », et construit son discours autour de relances syntaxiques de type « NN (nom) SENT (ponctuation forte) CC (conjonction de coordination) PP (pronom personnel) » (…*work. And I…*, …*Nation. And you…*, …*country. And we…* etc.) (figures 4 et 5)

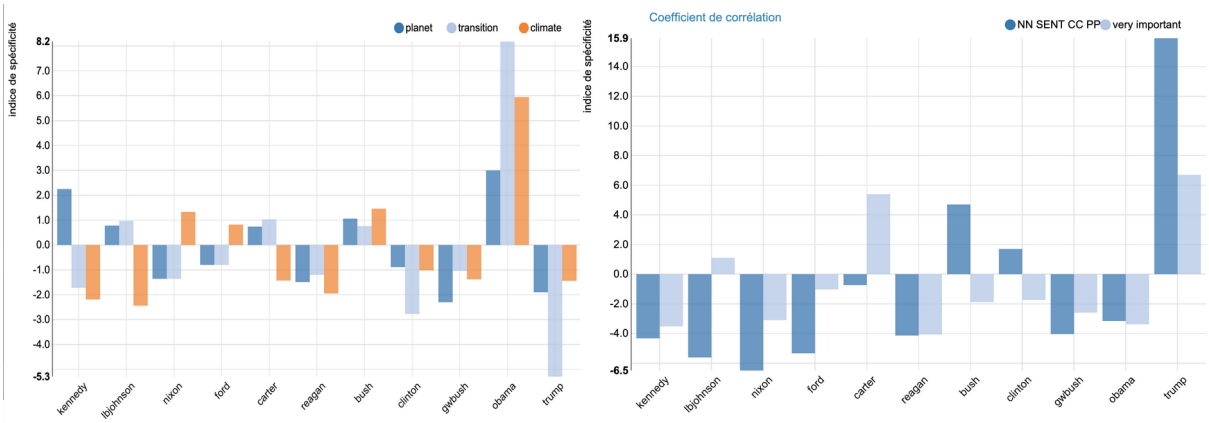

Figure 5 : Distribution statistique des marqueurs identifiés par le TDS dans le corpus anglais

Par hypothèse nous supposons que Macron, comme tout locuteur, construit son discours en empruntant consciemment ou inconsciemment à d'autres locuteurs (ici aux présidents du pays hôte). Le logiciel présenté repère ces emprunts ou ces empreintes. Il le fait tant d'un point de vue lexical (particulièrement lorsque les discours sont proches chronologiquement et que , en conséquence, les thématiques sont avoisinantes) que d'un point de vue grammatical (particulièrement lorsque les discours sont éloignés chronologiquement, que le lexique conjoncturel s'est évanoui, et que seules les permanences grammaticales demeurent).

4.2. Les Héroïdes *du poète latin Ovide, entre élégie et épopée*

Le poète latin Ovide, contemporain de l'empereur Auguste, a produit une œuvre littéraire abondante et variée, tantôt relevant de l'élégie amoureuse, tantôt portant sur des sujets à caractères mythologiques ou historiques. On ne s'étonnera donc pas de voir qu'Hyperdeep, entraîné sur un corpus de poètes latins classiques, n'attribue pas massivement l'œuvre d'Ovide à l'un de ceux-ci, mais la distribue assez largement entre ceux-ci. L'emprunte de deux auteurs se fait toutefois sentir plus massivement, celle de Properce en premier lieu (32,41 %), ensuite celle de Virgile (23,44 %) (Figure 6). La chose s'explique aisément quand on sait que la production du premier consiste essentiellement en des poésies amoureuses et celle du second en une vaste épopée mythologie centrée autour de l'histoire du héros troyen Enée. Des passages d'une seule et même œuvre sont toutefois perçus par Hyperdeep comme étant particulièrement représentatifs de Properce, d'autres de celle de Virgile. Il s'agit en l'occurrence de passages de lettres fictives appelées *Héroïdes* que diverses héroïnes de la mythologie gréco-romaine adressent à leurs amoureux qui les ont délaissées. L'oeuvre comporte donc bien à la fois une composante amoureuse et une dimension mythologique. Ce que l'on notera toutefois avec intérêt, c'est que l'attribution d'un passage soit à Properce, soit à Virgile repose sur des facteurs sensiblement différents. C'est sans étonnement que l'on notera que les passages attribués à Virgile le sont essentiellement sur base des lemmes (en vert).

Dans l'exemple de Virgile (Figure 6), on rencontre un grand nombre de lemmes spécifiques à l'oeuvre (Figure 7): la coloration virgilienne du texte d'Ovide repose ici essentiellement sur le contenu du texte et sur la thématique mythologique et épique (marquée par les personnages, Enée, Mézence, les Rutules ou par une action telle que *arma induere*, revêtir ses armes) . Toutefois, les séquences de lemmes jouent ici un rôle déterminant: MEZENTIVS_N INDVO ARMA se rencontre une fois chez Ovide , alors que chez Virgile ont trouve INDVO ARMA MEZENTIVS_N.

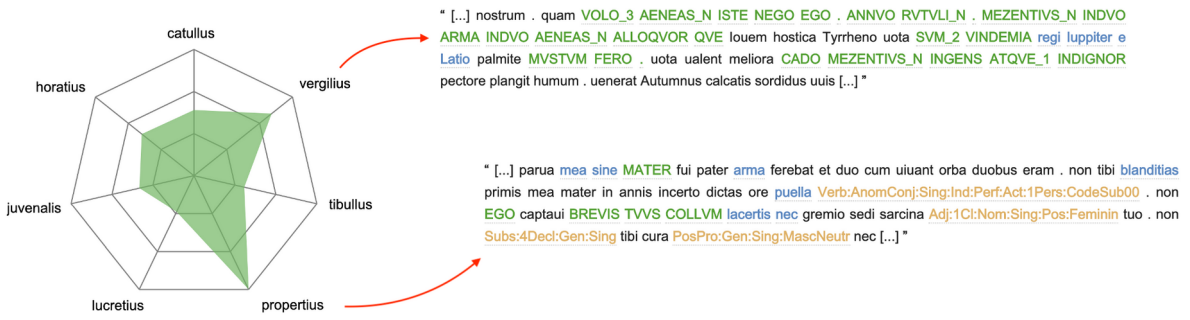

Figure 6 : Passages-clés d'Ovide attribués à Virgile et Properce.

Dans le cas de l'attribution à Properce de passages des *Héroïdes*, les codes grammaticaux semblent avoir un rôle plus important, comme le montre le deuxième exemple de la Figure 6, où la reconnaissance ne s'appuie pas seulement sur des lemmes (EGO: je) ou des formes (puella: jeune fille), liés à un échange amoureux, mais aussi sur 3 codes.

La spécificité de ces codes a très certainement ici joué une rôle, puisque ils ne sont spécifiques que de Properce, avec des excédents significatifs, alors que les trois sont en déficit chez Virgile (Figure 7).

Mais ici aussi la séquentialité des éléments a dû intervenir : en effet, les deux derniers des trois codes en orange dans le passage ci-dessus n'apparaissent séparés par 5 mots que chez Ovide et Properce. Le rôle joué tant par les codes que par leurs séquences s'explique ici fort bien : la parenté entre le deux textes ne repose pas seulement sur le contenu lexical des œuvres mais aussi sur diverses caractéristiques grammaticales propres à leur forme, comme l'emploi de formes de 1ère personne. Dans l'exemple d'Ovide, l'utilisation d'Hyperdeep permet d'objectiver ce que le philologue ne pouvait que pressentir à la lecture du texte.

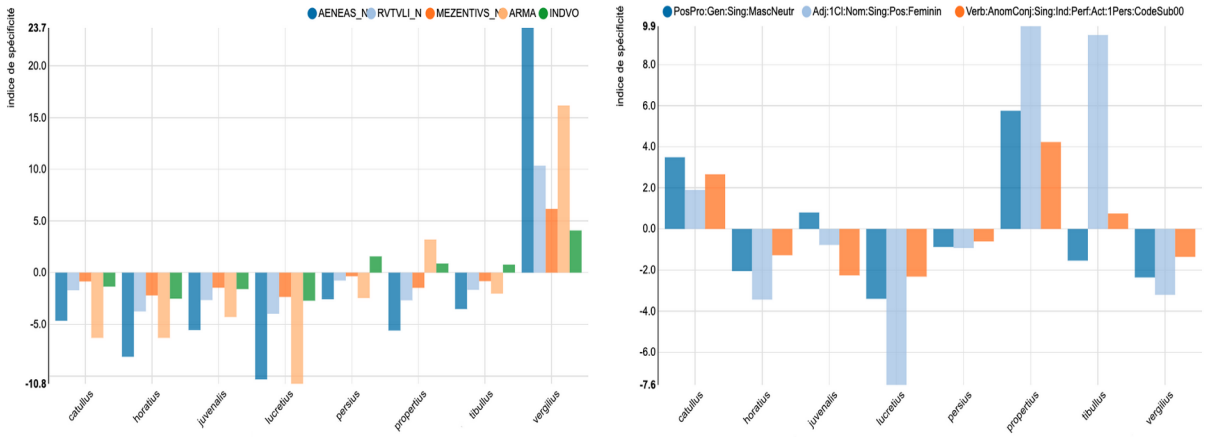

Figure 7 : Distribution statistique des marqueurs identifiés par le TDS dans le corpus latin

5. Conclusion et perspectives

Le deep learning semble avoir des qualités aussi bien prédictives que descriptives. Cette approche heuristique qui suit la même philosophie que la plupart des outils d'ADT donne un grain à moudre d'un genre nouveau aux chercheurs. La boite à outils statistiques jusqu'à présent disponible se voit donc enrichi avec Hyperdeep d'une nouvelle méthode, un nouveau regard sur les textes complémentaire des approches existantes. Le deep learning n'est plus boite noire, il est maintenant verbeux, très verbeux, peut-être trop à tel point que la statistique et l'expertise linguistique deviennent les alliés nécessaires pour une prise en main objective de ces nouveaux outils. C'est dans cette idée qu'Hyperdeep prolonge l'expérience proposée depuis plusieurs décennies par Hyperbase à travers une interface nouvelle qui fait le pari de marier statistique et deep learning dans un même outil.

Cependant les réseaux de neurones n'ont pas encore tous révélé leurs secrets. Par leur complexité accrue, les réseaux récurrent sont parmi les systèmes qui restent muets quant à leur mode de fonctionnement. Même si de nombreux travaux récents ont montré qu'il est possible d'analyser les couches cachées comme pour les réseaux convolutionnels, il reste à les interpréter et à les articuler avec l'existant. La modélisation de la langue, de la syntaxe ou du texte comme une séquence temporelle sont des pistes à creuser ces prochaines années. Les saillances textuelles identifiées par Hyperdeep devront être très prochainement complétées et comparées à celles des réseaux récurrents pour pouvoir enrichir encore un peu plus la méthode et l'ADT en général.

Bibliographie

Aghaebrahimian A., Cieliebak M. (2019). Hyperparameter tuning for deep learning in natural language processing. *4th Swiss Text Analytics Conference*, Winterthur.

Brunet E., Lebart L. et Vanni L. (2020 – sous presse). Littérature et intelligence artificielle. In Mayaffre D. et al. (2020 – sous presse), L'intelligence artificielle des textes. Points de vue critique, points de vue pratique. Honoré Champion.

Brunet E. et Vanni L. (2019). Deep learning et authentification des textes. Texto!, vol. XXIV - n°1 [http://www.revue-texto.net/docannexe/file/4194/texto_brunetvanni_deep_final.pdf], consulté le 06/01/2020].

Lebart L., Pincemin B. et Poudat C. (2019). Analyse des données textuelles. *Presses de l'Université du Québec*.

Lebart L. (1997). Réseaux de neurones et analyse des correspondances. In Modulad, (INRIA Paris), 18, pages 21–37.

Mayaffre D. et Vanni L. (2020). Objectiver l'intertexte ? Emmanuel Macron, deep learning et statistique textuelle. *15es Journées internationales d'Analyse statistique des Données Textuelles. JADT 2020* – soumis, Toulouse.

Montavon G., Samek W., Müller K.-R. (2018). Methods for Interpreting and Understanding Deep Neural Networks. *Digital Signal Processing*. 73. 1-15. 10.1016/j.dsp.2017.10.011.

Ribeiro M. T. , Singh S., Guestrin C. (2016). Why should i trust you? Explaining the predictions of any classifier. *In Proceedings of the 22nd international conference on knowledge discovery and data mining*, pages 1135–1144.

Strobelt H., Gehrmann S., Pfister H., Rush A. M. (2018). LSTMVis: A Tool for Visual Analysis of Hidden State Dynamics in Recurrent Neural *Networks. IEEE Transactions on Visualization and Computer Graphics* , 24 (1) , pp. 667-676.

Vanni L., Corneli M., Mayaffre D., Precioso F. (2020 – soumis) From text saliency to linguistic objects: learning linguistic interpretable markers with a multi-channels convolutional architecture, *In Proceedings Of The 58th Annual Meeting of the Association for Computational Linguistics.* ACL2020 – soumis, Seattle.

Vanni L., Ducoffe M., Mayaffre D., Precioso F., Longrée D. et al. (2018a). Textual Deconvolution Saliency (TDS) : a deep tool box for linguistic analysis *In Proceedings Of The 56th Annual Meeting of the Association for Computational Linguistics.* ACL 2018, Melbourne, Volume 1 p. 548-557.

Vanni L., Mayaffre D., Longrée D. (2018b). ADT et deep learning, regards croisés. Phrases-clefs, motifs et nouveaux observables. *14es Journées internationales d'Analyse statistique des Données Textuelles.* JADT 2018, Rome, p. 459-466.