
Réseaux de neurones de quaternions pour le traitement du langage

RÉSUMÉ. Les algorithmes d'apprentissage automatique, comme les réseaux de neurones (NN), ont permis d'atteindre des performances notables dans de nombreuses tâches liées au traitement automatique du langage (TAL). En TAL les contenus sont généralement représentés dans des espaces lexicaux ou thématiques. Les méthodes de traitement des contenus - y compris celles basées sur des NN - opèrent donc dans des espaces où les relations statistiques entre les éléments du document sont faiblement considérées. Nous proposons de remédier à cette faiblesse en étendant les réseaux de neurones aux quaternions (QMLP) pour permettre une modélisation des dépendances cachées entre les données. L'architecture proposée est évaluée lors d'une tâche d'analyse de dialogue parlés. Les résultats montrent que l'association du QMLP et d'une représentation structurée permet d'obtenir un gain en précision de 2% et 3% comparativement aux MLPs et à (Morchid et al., 2013) respectivement. Nous montrons finalement que la vitesse de convergence du QMLP est sensiblement supérieure à celle du MLP.

ABSTRACT. Machine Learning algorithms reach great performances on different Natural Language Processing tasks. Among these methods, Neural Networks (NN or MLP) recently received a great interest from researchers due to their capability to represent complex internal structures. However, MLPs employ basic word level or topic-based features and, therefore, reveal little in way of document statistical structure. We propose to address this issue by extending the NN to Quaternion called QMLP to take into consideration features dependencies. A well-dedicated segmentation of document approach is also compared to the one proposed in (Morchid et al., 2013). Experiments made on a SLU task with spoken dialogues show that our QMLP associated with the proposed document segmentation outperforms other approaches, with an gain of 2% and 3% compared to MLP and (Morchid et al., 2013) respectively. We finally demonstrated that less iterations are needed by QMLPs to reach better accuracies than MLP.

MOTS-CLÉS : Réseaux de neurones, Quaternions, Traitement du langage.

KEYWORDS: Neural Networks, Quaternions, Spoken Langage Understanding.

1. Introduction

Les méthodes fondées sur les quaternions se sont révélées extrêmement efficaces pour modéliser certaines propriétés telles que les rotations et le codage des images. Il existe de nombreuses applications des nombres hyper-complexes dans le domaine de l'intelligence artificielle telles que la vision assistée par ordinateur, le traitement de l'image, la robotique et récemment le traitement automatique du langage (Morchid *et al.*, 2013). Dans (Morchid *et al.*, 2013), des quaternions de caractéristiques de conversations agent-usager ont été proposés pour la première fois lors d'une tâche de traitement automatique du langage. Les conversations sont supposées évoluer en suivant un protocole prédéfini par l'agent. Les quaternions sont employés pour décrire les distributions des mots exprimant des thèmes potentiellement différents dans des segments de la conversation. Cependant, les conversations réelles ne suivent pas toujours un schéma conversationnel bien défini par des expressions ou positionné à un instant de l'interaction identique. Néanmoins, les quaternions de caractéristiques dépendants des segments proposés dans (Morchid *et al.*, 2013) ont montré leur pertinence à coder de l'information basées sur les segments de dialogue en comparaison à des distributions tenant compte de l'intégralité de la conversation. La méthode de classification employée dans (Morchid *et al.*, 2013) est basée sur le calcul de la distortion structurelle entre les représentations de deux documents. La classification était donc fortement dépendante du choix du document de référence, à partir duquel la rotation est définie.

Afin de résoudre ces difficultés, un perceptron multi-couche de quaternions (QMLP) est proposé comme classifieur. Le QMLP ainsi proposé est une nouvelle formulation appliquée au traitement automatique du langage de l'approche introduite par (Arena *et al.*, 1997). Notre QMLP permet d'outrepasser les problèmes suivants, et mentionnés par (Hinton, 2010) à propos des MLP basés sur les nombres réels :

- 1) Les dépendances statistiques entre les données entrantes (fréquences de mots ou la représentation en topics) ne sont pas complètement capturées par les réseaux de neurones basés sur les nombres réels.
- 2) La classification des thèmes basée sur des caractéristiques exprimées par des nombres réels est difficile pour un corpus contenant un grand nombre de documents pouvant contenir des sous-thèmes proche.
- 3) Les tâches comme la reconnaissance d'images ou la classification de documents sont représentées par des vecteurs d'entrée de grande taille qui ne sont pas bien caractérisés par les données non structurées que sont les nombres réels.

Les performances liées à la classification de conversations utilisant des QMLP varient, car dépendantes de la manière dont les données d'entrées ont été construites. Pour cette raison, cet article propose également d'investiguer différentes méthodes de pré-traitement pour comparer des méthodes de segmentation de document. Ces différents processus de segmentation montrent qu'un choix adapté de caractéristiques dépendantes du type de document (dialogue, document texte, ...) permettait d'améliorer la précision des QMLPs. De plus, l'algèbre des quaternions nous permet de

capturer les dépendances statistiques entre les données grâce au produit d'Hamilton (Hamilton, 1866). Cela est rendu possible par la multiplication de deux rotations, représentées par deux quaternions, suivant une géodésique autour d'une sphère dans l'espace \mathbb{R}^3 . De cette manière, les dépendances latentes sont reliées par le produit d'Hamilton aux structures statistiques des données latentes. Nous montrons également que le temps d'exécution est réduit grâce à un nombre d'itérations nécessaires à l'apprentissage de notre QMLP limité comparativement au MLP. De plus, le QMLP peut être utilisé quel que soit le type d'entrées afin d'améliorer les systèmes de traitement du langage (Arena *et al.*, 1997 ; Isokawa *et al.*, 2009). Notre proposition de QMLP a été évaluée dans le cadre de la tâche d'identification de thèmes du corpus du projet DECODA (Bechet *et al.*, 2012). Enfin, cette approche originale pourrait être utilisée dans un vaste spectre de domaines liés à l'intelligence artificielle, comme la vision assistée par ordinateur, ou l'étude et le traitement des mouvements (Aspragathos et Dimitros, 1998).

Ce papier est organisé de la façon suivante : L'algèbre des quaternions, l'architecture du QMLP, ainsi que les bases des MLP pour la compréhension de la différence avec notre proposition sont présentés dans les sections 2 et 3. La section 4 détaille le protocole expérimental tandis que les résultats obtenus sont reportés dans la section 5. Finalement, la section 6 conclut cette étude et expose les perspectives.

2. Description des données et des concepts de base des MLP

Un QMLP est ici proposé pour encoder les dépendances statistiques entre les données d'un document. Ces données sont utilisées pour identifier le thème dominant de ce document. Le terme "thème" est utilisé pour distinguer les sujets de conversation qui sont également les sorties de la classification à partir des données d'entrée obtenues avec des topics latents. Ces topics sont calculés en utilisant une "latent dirichlet allocation" (LDA), que nous allons maintenant brièvement décrire ainsi que les concepts basiques des MLP.

2.1. Représentation de la conversation dans un espace thématique LDA

Les caractéristiques latentes issues d'un espace thématique LDA (Blei *et al.*, 2003) ont montré leurs performances lors de nombreuses tâches telles que l'extraction de phrases (Bellegarda, 2000) ou de mots (Suzuki *et al.*, 1998). LDA est un modèle génératif probabiliste dans lequel les mots sont représentés par des mélanges de topics latents.

Certaines techniques telles que la méthode variationnelle (Griffiths et Steyvers, 2004), "Expectation-propagation" (Minka et Lafferty, 2002), ou *Gibbs Sampling* (Griffiths et Steyvers, 2004) détaillées dans (Heinrich, 2005), sont employées pour estimer les paramètres du modèle LDA ainsi que pour représenter un nouveau

document d dans l'espace de topics de taille T . La i^{ieme} caractéristique est calculée ainsi :

$$x_d^i = \theta_{(d,i)}, \quad [1]$$

où $\theta_{(d,i)} = P(z_i|d)$ est la probabilité que le topic z_i ($1 \leq i \leq T$) soit produit par le document d dans l'espace de thématique LDA de taille T .

2.2. Concepts de base des perceptrons multi-couches (MLP)

Considérons un MLP composé de M couches de neurones et x le vecteur d'entrée de ce MLP. Une fonction d'activation est utilisée pour calculer la valeur de sortie d'un noeud sachant son entrée x . Présentons à présent les opérations basiques pour obtenir une valeur de sortie d'un neurone sachant sa valeur d'entrée, ainsi que celles concernant l'estimation des paramètres du modèle pour un MLP basé sur les nombres réels avant d'introduire la version correspondante aux QMLP.

Fonction d'activation

La fonction d'activation utilisée durant les expérimentations est la fonction classique *sigmoid* (Duch et Jankowski, 1999) :

$$\alpha(x) = \frac{1}{1 + e^{-x}} \quad [2]$$

L'algorithme d'apprentissage des MLP est composé de 3 étapes : la propagation avant, la retro-propagation de l'erreur observée et la mise à jour des paramètres du modèle.

Propagation avant

Soit N_l le nombre de neurones contenus dans la couche l ($1 \leq l \leq M$), θ_n^l le biais du neurone n ($1 \leq n \leq N_l$) de la couche l . Sachant un ensemble P d'individus en entrée x_p ($1 \leq p \leq P$) et un ensemble de labels t_p associés à chaque x_p , la sortie γ_n^l ($\gamma_n^0 = x_p^n$) du neurone n de la couche l est :

$$\begin{aligned} \gamma_n^l &= \alpha(S_n^l) \\ \text{with } S_n^l &= \sum_{m=0}^{N_{l-1}} w_{nm}^l \times \gamma_m^{l-1} + \theta_n^l \end{aligned} \quad [3]$$

Retropropagation

L'erreur e observée entre la sortie espérée t et le résultat de la propagation avant γ est calculée pour la couche de sortie $l = M$ suivant :

$$e_n^l = t_n - \gamma_n^l \quad [4]$$

et pour la couche cachée ($1 \leq l < M$) :

$$e_n^l = \sum_{h=1}^{N_{l+1}} w_{h,n}^{l+1} \times \delta_h^{l+1}. \quad [5]$$

Le gradient δ est évalué par :

$$\delta_n^l = e_n^l \times \frac{\partial \alpha(S_n^l)}{\partial S_n^l} \text{ where } \frac{\partial \alpha(S_n^l)}{\partial S_n^l} = \alpha(S_n^l)(1 - \alpha(S_n^l)) \quad [6]$$

Adaptation des paramètres du modèle

Une fois les erreurs entre les sorties espérées et celles calculées lors de la phase de propagation déterminées, les poids $w_{n,m}^l$ ainsi que les biais θ_n^l sont respectivement mis-à-jour vers $\hat{w}_{n,m}^l$ et $\hat{\theta}_n^l$:

$$\hat{w}_{n,m}^l = w_{n,m}^l + \epsilon \delta_n^l \times \alpha(S_n^l) \quad [7]$$

$$\hat{\theta}_n^l = \theta_n^l + \epsilon \delta_n^l . \quad [8]$$

3. Perceptron Multicouche Quaternionique et segmentation des documents

La présente section décrit les bases de l'algèbre des quaternions, des perceptrons multicouche de quaternions (QMLP) ainsi que notre proposition de segmentation des documents.

3.1. Algèbre des quaternions

L'algèbre des quaternions \mathbb{Q} est une extension des nombres complexes définie dans un espace composé de quatre éléments dénotés $1, \mathbf{i}, \mathbf{j}, \mathbf{k}$ représentant une rotation. L'élément 1 correspond à l'identité. Un quaternion Q est défini de la façon suivante :

$$Q = r1 + x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \quad [9]$$

et représente une relation entre les quatre nombres réels r, x, y, z . Dans un quaternion, r est considéré comme la partie réelle tandis que $x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ forment la partie imaginaire (I) ou la partie vectorielle. Voici un ensemble de propriétés basiques sur les quaternions, importantes pour la définition des QMLP :

- Tous les produits possibles de \mathbf{i}, \mathbf{j} et \mathbf{k} :

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1 \quad [10]$$

- conjugué : Q^* of Q is : $Q^* = r1 - x\mathbf{i} - y\mathbf{j} - z\mathbf{k}$
- norme : $|Q| = \sqrt{r^2 + x^2 + y^2 + z^2}$
- quaternion normalisé : Q^∇

$$Q^\nabla = \frac{Q}{|Q|} \quad [11]$$

- Le produit scalaire entre deux quaternions $Q = r1 + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ et $Q' = r'1 + x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k}$ est :

$$\langle Q, Q' \rangle = rr' + xx' + yy' + zz' \quad [12]$$

• Le produit d'Hamilton \otimes entre $Q = r1 + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ et $Q' = r'1 + x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k}$ encode les dépendances latentes entre les caractéristiques. Celui-ci est défini par :

$$\begin{aligned} Q \otimes Q' = & (rr' - xx' - yy' - zz') + \\ & (rx' + xr' + yz' - zy')\mathbf{i} + \\ & (ry' - xz' + yr' + zx')\mathbf{j} + \\ & (rz' + xy' - yx' + zr')\mathbf{k} \end{aligned} \quad [13]$$

Ce produit permet d'effectuer une interpolation entre deux rotations suivant une géométrie autour d'une sphère dans l'espace \mathbb{R}^3 .

Sachant une segmentation $S = \{s_1, s_2, s_3, s_4\}$ d'un document $d \in D$ dépendant de la segmentation détaillée ci-après, ainsi qu'un ensemble de topics provenant de LDA $z = \{z_1, \dots, z_i, \dots, z_T\}$, chaque topic z_i d'un document d est représenté par le quaternion :

$$Q_d(z_i) = x_d^1(z_i)1 + x_d^2(z_i)i + x_d^3(z_i)j + x_d^4(z_i)k, \quad [14]$$

où $x_d^m(z_i)$ est la probabilité du topic z_i sachant le segment s_m d'un document d (voir Section 2.1). Ce quaternion est ensuite normalisé tel que décrit dans l'équation (11) afin d'obtenir l'entrée $Q_d^s(z_i)$ des QMLP.

Une approche plus approfondie des nombres hyper-complexes peut être trouvée dans (Kantor *et al.*, 1989 ; Kuipers, 1999 ; Zhang, 1997) et encore plus précisément à propos de leur algèbre dans (Ward, 1997).

3.2. Perceptron multicouche quaternionique (QMLP)

L'algorithme du QMLP est ici détaillé. Ce dernier diffère des MLP basés sur les nombres réels dans chaque sous processus d'apprentissage, et tous les éléments de la structure (entrées x , labels t , poids w , biais b , sorties γ, \dots) sont des quaternions :

Fonction d'activation

La fonction d'activation β est composée de la fonction *sigmoid* α , définie par l'équation (2), appliquée à chaque élément du quaternion $Q = r1 + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ suivant la proposition de (Arena *et al.*, 1997) :

$$\beta(x) = \alpha(r)1 + \alpha(x)\mathbf{i} + \alpha(y)\mathbf{j} + \alpha(z)\mathbf{k} \quad [15]$$

Propagation avant

Tout comme pour le MLP, N_l est le nombre de neurones contenus dans la couche l ($1 \leq l \leq M$) et M le nombre total de couches du QMLP. θ_n^l est le biais du neurone n ($1 \leq n \leq N_l$) de la couche l . Sachant un ensemble P d'individus (quaternions)

d'entrée x_p ($1 \leq p \leq P$) et un ensemble de labels t_p associés à chacun de ces x_p , la sortie γ_n^l ($\gamma_n^0 = x_p^n$) du neurone n de la couche l est défini par :

$$\begin{aligned} \gamma_n^l &= \beta(S_n^l) \\ \text{with } S_n^l &= \sum_{m=0}^{N_{l-1}} w_{nm}^l \otimes \gamma_m^{l-1} + \theta_n^l \end{aligned} \quad [16]$$

Retropropagation

L'erreur e observée entre la sortie espérée y et le résultat de la propagation avant γ est ensuite évaluée pour la couche de sortie $l = M$ suivant :

$$e_n^l = t_n - \gamma_n^l, \quad [17]$$

et pour la couche cachée ($1 \leq l < M$) :

$$e_n^l = \sum_{h=1}^{N_{l+1}} w_{h,n}^{*l+1} \otimes \delta_h^{l+1}, \quad [18]$$

Le gradient δ est calculé avec :

$$\delta_n^l = e_n^l \times \frac{\partial \beta(S_n^l)}{\partial S_n^l} \text{ where } \frac{\partial \beta(S_n^l)}{\partial S_n^l} = \beta(S_n^l)(1 - \beta(S_n^l)) \quad [19]$$

Adaptation des paramètres du modèle

Une fois les erreurs entre les sorties espérées et les résultats calculées, les poids $w_{n,m}^l$ et les biais θ_n^l sont adaptés respectivement en $\hat{w}_{n,m}^l$ et $\hat{\theta}_n^l$:

$$\hat{w}_{n,m}^l = w_{n,m}^l + \epsilon \delta_n^l \otimes \beta^*(S_n^l) \quad [20]$$

$$\hat{\theta}_n^l = \theta_n^l + \epsilon \delta_n^l. \quad [21]$$

3.3. Segmentation du document

Une première segmentation simple, nommée par la suite **SEG_1**, a été introduite par (Morchid *et al.*, 2013). Cette dernière considère 4 segments consécutifs d'approximativement la même taille excluant toute ponctuation et délimiteurs de tour de parole (découpage possible au sein d'une même phrase). La représentation du document en fonction de ces 4 segments est uniquement utilisée comme *baseline* afin de mesurer les performances obtenues à l'aide d'une segmentation plus adaptée à la structure du dialogue appelée **SEG-UA** pour segmentation Usager-Agent. Ce modèle est donc basé sur la structure interne du dialogue et considère séparément les caractéristiques de l'usager (**i**), de l'agent (**j**) ainsi que de l'ensemble du document (**k**).

4. Protocol expérimental

4.1. Corpus de dialogues parlés

Le corpus de documents est un ensemble de conversations téléphoniques automatiquement traduites par un système de reconnaissance automatique de la parole (SRAP), provenant du service-client de la société de transport de PARIS (RATP). Ce corpus provient du projet DECODA (Bechet *et al.*, 2012) et est utilisé pour évaluer la pertinence de notre segmentation fondée sur la structure du dialogue ainsi que sur une représentation en quaternion de ces caractéristiques extraites dans un QMLP lors d’une tâche d’identification des thèmes. Le corpus DECODA est composé de 1 242 conversations téléphoniques, correspondant à approximativement 74 heures de signal de parole. Le corpus est découpé en 8 thèmes suivant la répartition décrite par le Tableau 1.

Tableau 1. *Corpus DECODA.*

Class labels	Nombre d'individus		
	apprentissage	développement	test
problèmes d'itinéraires	145	44	67
objets trouvés / perdus	143	33	63
horraires	47	7	18
cartes de transport	106	24	47
état du trafic	202	45	90
amendes	19	9	11
infractions	47	4	18
offres spéciales	31	9	13
Total	740	175	327

Le système de reconnaissance automatique de la parole (SRAP) LIA-Speeral (Linares *et al.*, 2007) a été utilisé pour extraire le contenu textuel des dialogues du corpus DECODA. Les paramètres du modèle acoustique ont été estimés sur 150 heures de conversations téléphoniques. Le vocabulaire contient 5 782 mots. Un modèle de langage 3-gram a été obtenu en adaptant un modèle de langage basique avec les transcriptions du corpus d’entraînement. Une “stop list” de 126 mots¹ a été utilisée pour retirer les mots inutiles, permettant ainsi au taux d’erreur mot d’atteindre 33.8% sur l’entraînement, 45.2% sur le développement et 49.5% sur le test. Ces taux d’erreurs élevés sont principalement due aux disfluences dans le signal de parole ainsi qu’aux conditions difficiles de l’environnement acoustique (par exemple, le bruit d’une rue à travers un téléphone mobile).

1. <http://code.google.com/p/stop-words/>

4.2. Espaces de topics LDA et configurations des MLP

Différentes configurations d’espaces LDA pour les deux types de vecteurs d’entrées (Document entier pour les réels, document segmentés pour les quaternions) sont utilisées.

Configuration des espaces de topics LDA

Différents modèles LDA ont été appris en variant le nombre de topics T de 5 à 100 par pas de 1. Les modèles LDA requièrent également de choisir les hyper-paramètres α et β qui contrôlent respectivement la distribution du topic dans le document et la distribution des mots dans les topics. L’heuristique standard est $\alpha = \frac{50}{T}$ et $\beta = 0.01$ (Griffiths et Steyvers, 2004).

Configurations des MLP et QMLP

Les expérimentations comparent 3 réseaux de neurones dont deux basés sur des réels (MLP, MLP4) et un sur les nombres hyper-complexes (QMLP). Chaque réseau de neurones comporte une couche cachée composée de 8 noeuds :

1) Un **MLP** classique avec un ensemble T de caractéristiques ($T =$ nombre de topics dans l’espace LDA) pour vecteur d’entrée et 8 sorties correspondantes au nombre de thèmes contenus dans le corpus DECODA.

2) Notre architecture **QMLP** basée sur les entrées provenant de vecteurs de dimension T de quaternions composés de la probabilité de chaque topic pour : 1) **SEG_1** chacune des quatre parties du dialogue est projetée dans l’espace thématique, et la probabilité de cette partie est ensuite portée dans le quaternion (par exemple, la probabilité du premier topic LDA sachant le premier quart du dialogue est insérer dans la partie (r) du premier quaternion); 2) **SEG-UA** la partie usager ($x\mathbf{i}$), agent ($y\mathbf{j}$) et le document entier ($z\mathbf{k}$) afin de respectivement composer la partie imaginaire du quaternion (méthode également utilisée dans le traitement de l’image pour le spectre $R(x\mathbf{i})G(y\mathbf{j})B(z\mathbf{k})$). Tandis que la partie réelle r est initialisée à 0. Durant la phase d’apprentissage, toutes les valeurs du quaternion de sortie correspondant au thème annoté sont mises à 1, tandis que tous les coefficients des autres quaternions sont initialisés à 0.

3) Un MLP dont le vecteur d’entrée est une concaténation des trois partie **ijk** des quaternions appelé **MLP4** est également proposé pour une comparaison équitable avec le QMLP. Cette méthode permettra également de mesurer l’impact du produit hamiltonien durant l’apprentissage.

5. Results

Les résultats reportés dans la Figure 1 proviennent d’un “k-fold” ($k = 10$) afin d’obtenir une comparaison robuste et convaincante entre les différents processus de segmentation ainsi que entre les différents MLP (valeurs réels et de quaternions). La précision obtenue durant la tâche d’identification des thèmes de dialogues parlés à partir des MLP, MLP4, et QMLP est donc reportée par la Figure 1 utilisant respectivement les méthodes de segmentation **SEG_1** (a) et (b) , et **SEG-UA** (c) et (d). La

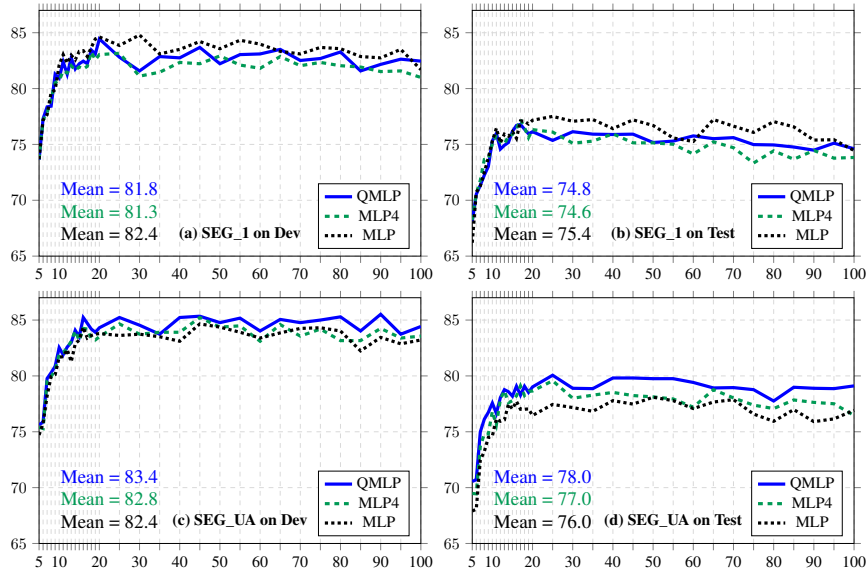


Figure 1. Précision en % obtenue en faisant varier le nombre de topics T du modèle LDA pour le MLP, MLP4 et le QMLP avec les segmentations *SEG_1* (a)-(b) et *SEG-UA* (c)-(d)

Table 3 résume elle, les différentes précisions obtenues à partir de *SEG_1* tandis que la Table 4 résume les résultats observés à partir de la segmentation adaptée du dialogue *SEG-UA* introduite dans ce papier.

5.1. Impact de la segmentation du document

Deux segmentations différentes du document ont été proposées afin de construire des vecteurs d'entrée du QMLP : 1) *SEG_1* est une méthode de segmentation simpliste basée sur le nombre de mots contenus dans le dialogue sans considération aucune de la structure de ce dialogue (Morchid *et al.*, 2013). Cette représentation est basée sur la probabilité du topic pour chacune des 4 parties égaux en nombre de mots du document ; 2) *SEG-UA* est une proposition plus appropriée au regard de la structure interne du dialogue. Le quaternion représente ainsi la probabilité d'un topic du modèle LDA pour la partie usager (x_i), agent (y_j) et enfin du document entier afin d'extraire le thème général de ce document (z_k) (représentation thématique adapté à l'utilisateur, l'agent et sur l'ensemble du document).

La première remarque est que les résultats obtenus avec notre proposition de méthode de segmentation basée sur la structure des dialogues (*SEG-UA*), reportés dans le tableau 2, surpassent ceux observés à l'aide de la méthode de segmentation plus

Modèle	Type de segmentation	Moyenne sur l'ensemble de test	Amélioration
(Morchid <i>et al.</i> , 2013)	SEG_1	73.9	–
MLP	SEG_1	75.46	–
MLP4	SEG_1	74.63	–
QMLP	SEG_1	74.84	–
MLP	SEG-UA	76.01	+0.55
MLP4	SEG-UA	77.06	+2.43
QMLP	SEG-UA	78.05	+3.21

Tableau 2. Précision en % obtenue durant la tâche d'identification de thèmes de dialogues parlés à partir des segmentations *SEG_1* et *SEG-UA*

classique (*SEG_1*) quelque soit le réseau de neurones employé, avec un gain de 2 et 3 points (la dernière colonne du tableau2) pour le MLP4 et QMLP respectivement. La similitude des résultats reportés pour le MLP sur les deux représentations de document est due au processus d'initialisation ainsi que les 10 échantillons employés, mais sont assez équivalents (moins de 0.6% de différence). De plus, le gain observé pour le QMLP (3.2 points) est plus important que celui obtenu avec le MLP4 (2.4 points). Ceci est principalement dû au fait que les coefficients imaginaires d'un même quaternion encodent les dépendances des caractéristiques de *SEG-UA* qui ne sont pas capturées par le MLP4. Il est également important de souligner que les précisions produites par les modèles MLP et QMLP sont plus robuste avec la segmentation *SEG-UA* comme reporté dans la Figure 1. En effet, les précisions décrites par les courbes (a) et (b) pour *SEG_1* décroissent pour des modèles LDA contenant un nombre de topics T supérieur à 70. Ce phénomène n'apparaît pas lorsque la méthode *SEG-UA* est employée (courbes (c) et (d) dans la Figure 1). Nous avons également séparément évalué les représentations des usagers, des agents, et ainsi obtenu des précisions maximales pour le test de 66% et 73.5% respectivement. Pour aller plus loin, à partir des données de *SEG-UA* ainsi qu'un nombre de topics compris entre 40 et 45, de très bonnes précisions sont observées autant avec le corpus de DEV qu'avec le corpus de TEST. Cela permet donc d'utiliser les résultats du DEV pour prédire le nombre de topics pour lequel la précision maximale sur le TEST est atteinte. Ces résultats confirment que des caractéristiques séparées pour l'agent sont efficaces pour caractériser le fait que la tâche de l'agent est d'orienter la conversation vers un thème spécifique.

5.2. QMLP vs. MLP

Il est important de souligner que le QMLP obtient de meilleures performances avec *SEG-UA* que le MLP ainsi que le MLP4 avec un gain de 2 et 1 points respectivement pour la moyenne du TEST entre tous les modèles LDA (T allant de 5 à 100), tel que présenté dans le Tableau 4. Cette remarque est également vérifiée pour la meilleure précision sur l'ensemble de développement, ainsi que la précision la plus élevée sur l'ensemble de test.

Modèle	Max dev.	Max test	Moy. test	#Iterations
(Morchid <i>et al.</i> , 2013)	82.2	73.9	73.9	–
MLP	84.79	77.50	75.46	965.5
MLP4	83.27	77.23	74.63	154.2
QMLP	84.44	76.79	74.83	81.2

Tableau 3. Précision en % obtenue durant la tâche d’identification à partir de la segmentation *SEG_1*

L’intuition initiatrice de ce papier que le QMLP obtiendrait de meilleurs résultats lorsque les caractéristiques extraites du document seraient adaptées à la représentation en quaternions est vérifiée. En effet, le Tableau 3 et la Figure 1 (a)-(b) montrent que le QMLP parvient difficilement à des précisions équivalentes aux MLP et MLP4 en se basant sur une segmentation de documents non pertinente et ne révélant que peu de dépendances entre les caractéristiques d’un document (*SEG_1*). Il est clair que le nombre d’itérations nécessaire à l’apprentissage du modèle QMLP est largement inférieur à celui des autres modèles (MLP et MLP4). Par exemple en comparaison avec le MLP, le QMLP nécessite 7 fois moins d’itérations pour la segmentation *SEG-UA* et 12 fois moins pour la segmentation *SEG_1*. Bien que le temps d’exécution total soit un peu plus long pour la segmentation *SEG-UA* du QMLP, cette représentation permet d’obtenir de meilleurs performances.

Modèle	Max dev.	Max test	Moy. test	#Iterations
MLP	84.65	78.09	76.01	882.9
MLP4	85.23	79.57	77.06	285.3
QMLP	85.51	80.06	78.05	122.3

Tableau 4. Précision en % obtenue durant la tâche d’identification à partir de la segmentation *SEG-UA*

Globalement, le QMLP proposé met en lumière les dépendances statistiques latentes lorsque la structure interne du document est bien décrite lors de la construction des vecteurs d’entrées de quaternions. Ceci est dû au fait que l’algorithmique des quaternions révèle l’interpolation entre quaternions et donc entre les topics d’un document à l’aide de la multiplication d’Hamilton.

6. Conclusion

Nous avons proposé une approche originale basée sur l’algèbre des quaternions et les réseaux de neurones pour le traitement de document parlés dans un contexte applicatif réel. Cette représentation tire un avantage de la bonne segmentation des documents tandis que le MLP ne parvient pas à extraire les structures internes sous-jacentes.

Les perceptrons multicouches basés sur les nombres réels et sur les quaternions sont ici tout deux évalués durant une tâche commune d'identification de thèmes pour des dialogues parlés. Les résultats obtenus démontrent que notre réseau de neurones de quaternions (QMLP) réduit considérablement le nombre d'itérations nécessaires lors de la phase d'apprentissage tout en obtenant de meilleurs résultats. De plus, les intuitions initiales suivantes ont été démontrées : 1) La segmentation des documents est cruciale pour construire des caractéristiques robustes et 2) le produit d'Hamilton permet aux QMLP de révéler les dépendances statistiques entre les topics.

Dans des travaux futur, nous proposerons une implémentation sur GPU d'un QMLP afin de minimiser le temps de traitement (CPU) ainsi que le temps de calcul (GPU). Des réseaux de neurones de quaternions profonds seront également investigués afin de totalement révéler le potentiel des structures profondes. Comme montré précédemment, la segmentation des documents est cruciale. De ce fait, nous explorerons ensuite différentes représentations du document basées sur les topics ainsi que sur les mots pour révéler les informations et relations latentes. Pour aller plus loin, il serait intéressant de comparer notre QMLP aux réseaux de neurones convolutionnels et récurrents décrits par (Krizhevsky *et al.*, 2012) et (Sak *et al.*, 2015) qui se sont récemment montrés très efficaces pour capturer les observations latentes séquentielles.

7. Bibliographie

- Arena P., Fortuna L., Muscato G., Xibilia M. G., « Multilayer perceptrons to approximate quaternion valued functions », *Neural Networks*, vol. 10, n° 2, p. 335-342, 1997.
- Aspragathos N. A., Dimitros J. K., « A comparative study of three methods for robot kinematics », *Systems, Man, and Cybernetics, Part B : Cybernetics, IEEE Transactions on*, vol. 28, n° 2, p. 135-145, 1998.
- Bechet F., Maza B., Bigouroux N., Bazillon T., El-Beze M., De Mori R., Arbillot E., « DECODA : a call-centre human-human spoken conversation corpus. », *LREC*, p. 1343-1347, 2012.
- Bellegarda J. R., « Exploiting latent semantic information in statistical language modeling », *Proceedings of the IEEE*, vol. 88, n° 8, p. 1279-1296, 2000.
- Blei D. M., Ng A. Y., Jordan M. I., « Latent dirichlet allocation », *the Journal of machine Learning research*, vol. 3, p. 993-1022, 2003.
- Duch W., Jankowski N., « Survey of neural transfer functions », *Neural Computing Surveys*, vol. 2, n° 1, p. 163-212, 1999.
- Griffiths T. L., Steyvers M., « Finding scientific topics », *Proceedings of the National academy of Sciences of the United States of America*, vol. 101, n° 1, p. 5228-5235, 2004.
- Hamilton S., *Elements of quaternions*, Longmans, Green, & co., 1866.
- Heinrich G., « Parameter estimation for text analysis », *Web : [http ://www. arbylon. net/publications/text-est. pdf](http://www.arbylon.net/publications/text-est.pdf)*, 2005.
- Hinton G., « A practical guide to training restricted Boltzmann machines », *Momentum*, vol. 9, n° 1, p. 926, 2010.

- Isokawa T., Matsui N., Nishimura H., « Quaternionic neural networks : Fundamental properties and applications », *Complex-Valued Neural Networks : Utilizing High-Dimensional Parameters*, vol. 10, p. 411-439, 2009.
- Kantor I., Solodovnikov A., Shenitzer A., *Hypercomplex numbers : an elementary introduction to algebras*, Springer-Verlag, 1989.
- Krizhevsky A., Sutskever I., Hinton G. E., « Imagenet classification with deep convolutional neural networks », *Advances in neural information processing systems*, p. 1097-1105, 2012.
- Kuipers J. B., *Quaternions and rotation sequences*, Princeton university press Princeton, NJ, USA :, 1999.
- Linares G., Nocéra P., Massonie D., Matrouf D., « The lia speech recognition system : from 10xrt to 1xrt », *Text, Speech and Dialogue*, Springer, p. 302-308, 2007.
- Minka T., Lafferty J., « Expectation-propagation for the generative aspect model », *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*, Morgan Kaufmann Publishers Inc., p. 352-359, 2002.
- Morchid M., Linarès G., El-Beze M., De Mori R., « Theme Identification in Telephone Service Conversations using Quaternions of Speech Features », *Interspeech*, ISCA, 2013.
- Sak H., Senior A., Rao K., Beaufays F., « Fast and accurate recurrent neural network acoustic models for speech recognition », *arXiv preprint arXiv :1507.06947*, 2015.
- Suzuki Y., Fukumoto F., Sekiguchi Y., « Keyword extraction using term-domain interdependence for dictation of radio news », *17th international conference on Computational linguistics*, vol. 2, ACL, p. 1272-1276, 1998.
- Ward J., *Quaternions and Cayley numbers : Algebra and applications*, vol. 403, Springer, 1997.
- Zhang F., « Quaternions and matrices of quaternions », *Linear algebra and its applications*, vol. 251, p. 21-57, 1997.