

QUATERNION NEURAL NETWORKS FOR SPOKEN LANGUAGE UNDERSTANDING

*Titouan Parcollet[†], Mohamed Morchid[†], Pierre-Michel Bousquet[†], Richard Dufour[†]
Georges Linarès[†] and Renato De Mori[‡], Fellow, IEEE*

[†]LIA, University of Avignon (France)

[‡]McGill University, School of Computer Science, Montreal, Quebec (Canada)

{firstname.lastname}@univ-avignon.fr, rdemori@cs.mcgill.ca

ABSTRACT

Machine Learning (ML) techniques have allowed a great performance improvement of different challenging Spoken Language Understanding (SLU) tasks. Among these methods, Neural Networks (NN), or Multilayer Perceptron (MLP), recently received a great interest from researchers due to their representation capability of complex internal structures in a low dimensional subspace. However, MLPs employ document representations based on basic word level or topic-based features. Therefore, these basic representations reveal little in way of document statistical structure by only considering words or topics contained in the document as a “bag-of-words”, ignoring relations between them. We propose to remedy this weakness by extending the complex features based on Quaternion algebra presented in [1] to neural networks called QMLP. This original QMLP approach is based on hyper-complex algebra to take into consideration features dependencies in documents. New document features, based on the document structure itself, used as input of the QMLP, are also investigated in this paper, in comparison to those initially proposed in [1].

Experiments made on a SLU task from a real framework of human spoken dialogues showed that our QMLP approach associated with the proposed document features outperforms other approaches, with an accuracy gain of 2% with respect to the MLP based on real numbers and more than 3% with respect to the first Quaternion-based features proposed in [1]. We finally demonstrated that less iterations are needed by our QMLP architecture to be efficient and to reach promising accuracies.

Index Terms— Quaternion, Neural Network, Spoken Language Understanding, Machine Learning

1. INTRODUCTION

Quaternions are number systems that have been found very useful to model properties such as rotation that are described

by a number system. They have found applications in artificial intelligence domains such as computer graphics, computer vision and recently on spoken language processing [1]. In [1], quaternion features for agent-customer conversations have been considered for the first time for an NLP task. As conversations are supposed to evolve according to a protocol followed by the agent, quaternions are used to describe word distributions expected to express possibly different topic clues in conversation segments. Real-life conversations not always follow a conversation model with expression of specific contents in specific segments. Nevertheless, segment dependent quaternions used in [1] have shown to provide interesting topic features compared to whole conversation word distributions. The type of classification used in [1] is based on computing the structural distortion between the representations of two documents. The classification accuracy was highly dependent on the choice of the reference documents with which rotation was computed. In order to avoid such a difficulty, a Quaternion Multi-Layer Perceptron (QMLP) is proposed as a classifier. The proposed QMLP is a new formulation applied to natural language processing (NLP) of an approach to quaternion function approximation introduced in [2]. Our QMLP makes it possible to alleviate the following problems mentioned in [3] about the use of just real numbers in MLP classifiers:

1. Statistical dependencies between input features (word frequencies or topic-based features) are not completely captured by real-valued NNs.
2. Topic classification based on real number features is difficult for large amounts of documents containing different close sub-topics.
3. Tasks such as image recognition or document classification have large inputs that are not well characterized by unstructured real number features.

Conversation classification performance using QMLPs is expected to exhibit variations depending on how input quaternion features are formed. For this reason, this paper also proposes to investigate different pre-processing methods to com-

This work was funded by the GAFES project supported by the French National Research Agency (ANR) under contract ANR-14-CE24-0022.

pare for segmenting documents. These different segmentation processes show that an adapted choice of features depending on the document type (dialogues, textual documents, ...) makes it possible to improve QMLP accuracies. Furthermore, quaternion algebra allows us to capture features statistical dependencies with the Hamilton product [4]. This is due to the multiplication of two rotations, represented by two quaternions, following a geodesic over a sphere in the \mathbb{R}^3 space. In this way, latent dependencies are related by the Hamilton product with the statistical structure of latent features. The processing time is also reduced thanks to a limited iterations number required to learn the QMLP models. Therefore, Quaternion MLP may be used with any kind of inputs to improve NLP systems [2, 5]. The proposed QMLP is evaluated with the theme identification task of the DECODA corpus [6]. Moreover, the original proposed approach can be employed in a broad spectrum of artificial intelligence domains such as computer vision, motion tracking and management [7].

The reminder of the paper is organized as follows: The proposed Quaternion's algebra, the QMLP architecture, as well as the basic concepts of MLPs to understand the difference with the proposed QMLP are presented in Sections 2 and Section 3. Section 4 details the experimental protocol while the results obtained are reported in Section 5. Finally, Section 6 concludes this study and gives some words to the perspectives.

2. DESCRIPTION OF DOCUMENT FEATURES AND MULTI-LAYER PERCEPTRON CONCEPTS

A QMLP is proposed in this paper to encode statistical dependencies between features of a document. These features are used for identifying the dominant theme of the document. The term "theme" is used to distinguish the conversation topics that are the classification outputs from input features obtained with latent topics. These topics are computed with latent Dirichlet allocation (LDA) that is now briefly reviewed as well as the basic concepts of MLP.

2.1. Topic-based features from a LDA

Features based on LDA [8] have demonstrated to be useful in various tasks such as sentences [9] or keywords [10] extraction. LDA is the basis of a generative model in which words are represented by mixture probabilities of latent topics.

Several techniques, such as Variational Methods [8], Expectation-propagation [11] or Gibbs Sampling [12], have been proposed to estimate the parameters describing the LDA hidden space. The Gibbs Sampling, reported in [12] and detailed in [13], is used to estimate LDA parameters and to represent a new document d in a topic space of size T . This model extracts a set of features from the topic-based representation of d . The i^{th} feature is computed as follows:

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

where $\theta_{(d,i)} = P(z_i|d)$ is the probability of topic z_i ($1 \leq i \leq T$) to be produced by the document d in the LDA topic space of size T .

2.2. Basic Concepts of a Multi-Layer Perceptron (MLP)

Let us consider a MLP made of M layers of nodes whose number depends on the layer. Let x be the input to a node. This input is obtained by adding a bias to the convolution of the outputs of the previous layer with a vector of model weights. An activation function is used for computing the output value of a node given its input. Basic operations for obtaining a node output from the node input and for estimating model parameters are now reviewed for real valued MLPs before introducing the corresponding QMLP version.

Activation function

The activation function used during the experiments is the classical *sigmoid* function [14]:

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

The feed-forward algorithm of the MLP is composed of 3 steps: forward, learning and update phases:

Forward phase

Let N_l be the number of neurons contained into the layer l ($1 \leq l \leq M$). θ_n^l is the bias of the neuron n ($1 \leq n \leq N_l$) from the layer l . Given a set of P input patterns x_p ($1 \leq p \leq P$) and a set of labels t_p associated to each x_p , the output γ_n^l ($\gamma_n^0 = x_p^n$) of the neuron n from the layer l is given by:

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

Learning phase

The error e observed between the expected outcome t and the result of the forward phase γ is then evaluated for the output layer $l = M$ as follows:

$$e_n^l = t_n - \gamma_n^l \quad (4)$$

and for the hidden layer ($1 \leq l < M$):

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

The gradient δ is computed with:

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

Update phase

When errors between the expected outcome and the result are computed, the weights $w_{n,m}^l$ and the bias θ_n^l have to be respectively updated to $\hat{w}_{n,m}^l$ and $\hat{\theta}_n^l$:

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

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

3. PROPOSED QUATERNION MULTI-LAYER PERCEPTRON AND DOCUMENT SEGMENTATION

This section describes the basic concepts of Quaternion algebra and Quaternion Multi-Layer Perceptron (QMLP) as well as the proposed well-tailored document segmentation.

3.1. Quaternion algebra

Quaternion algebra \mathbb{Q} is an extension of a complex number uniquely defined in a four dimensional space as a linear combination of four basis elements denoted as $1, \mathbf{i}, \mathbf{j}, \mathbf{k}$ to represent a rotation. The element 1 is the identity element of the vector space. A quaternion Q is written as:

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

and represents a relation between the four real numbers r, x, y, z . In a quaternion, r is its real part while $x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ is the imaginary part (I) or the vector part. There is a set of basic Quaternion properties that are important for the further QMLP definition:

- all the possible products of \mathbf{i}, \mathbf{j} and \mathbf{k} :

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

- quaternion conjugate Q^* of Q is: $Q^* = r1 - x\mathbf{i} - y\mathbf{j} - z\mathbf{k}$
- quaternion norm: $|Q| = \sqrt{r^2 + x^2 + y^2 + z^2}$
- normalized quaternion Q^\triangleleft

$$Q^\triangleleft = \frac{Q}{|Q|} \quad (11)$$

- inner product between two quaternions $Q = r1 + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ and $Q' = r'1 + x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k}$ is:

$$\langle Q, Q' \rangle = rr' + xx' + yy' + zz' \quad (12)$$

- Hamilton product \otimes between $Q = r1 + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ and $Q' = r'1 + x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k}$ encodes latent dependencies between latent features and is defined as follows:

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

It performs an interpolation between two rotations following a geodesic over a sphere in the \mathbb{R}^3 space.

Given a segmentation $S = \{s_1, s_2, s_3, s_4\}$ of a document $d \in D$ depending on the document segmentation detailed further and a set of topics from a LDA $z = \{z_1, \dots, z_i, \dots, z_T\}$, each topic z_i in a document d is represented by the quaternion:

$$Q_d(z_i) = x_d^1(z_i)1 + x_d^2(z_i)\mathbf{i} + x_d^3(z_i)\mathbf{j} + x_d^4(z_i)\mathbf{k} \quad (14)$$

where $x_d^m(z_i)$ is the prior of the topic z_i in segment s_m of a document d as described in Section 2.1. This quaternion is then normalized based on equation (11) to obtain the input $Q_d^\triangleleft(z_i)$ of QMLPs.

More about hyper-complex numbers systems can be found in [15, 16, 17] and more precisely about quaternion in [18].

3.2. Quaternion Multi-Layer Perceptron (QMLP)

This section details the QMLP algorithm. The QMLP differs from the real numbers MLP in each learning subprocess and all elements of the structure (input x , labels t , weights w , biases b , neuron outputs γ, \dots) are quaternions:

Activation function

The activation function β is composed with the *sigmoid* function α , defined in equation (2), applied to each element of the quaternion $Q = r1 + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ as follow [2]:

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

Forward phase

As for MLP, let N_l be the number of neurons contained in the layer l ($1 \leq l \leq M$) and M be the number of layers of the QMLP. θ_n^l is the bias of the neuron n ($1 \leq n \leq N_l$) from the layer l . Given a set of P quaternion input patterns x_p ($1 \leq p \leq P$) and a set of labels t_p associated to each x_p , the output γ_n^l ($\gamma_n^0 = x_p^n$) of the neuron n of the layer l is given by:

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

Learning phase

The error e observed between the expected outcome y and the result of the forward phase γ is then evaluated for the output layer $l = M$ as follows:

$$e_n^l = t_n - \gamma_n^l \quad (17)$$

and for the hidden layer ($1 \leq l < M$):

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

The gradient δ is computed with:

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

Update phase

When errors between the expected outcome and the result are computed, the weights $w_{n,m}^l$ and the bias values θ_n^l have to be respectively updated to $\hat{w}_{n,m}^l$ and $\hat{\theta}_n^l$:

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

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

3.3. Document segmentation

A straightforward quaternion-based segmentation, referred to later on in this paper as **SEG-1**, has been proposed in [1]. Such a segmentation considers 4 successive segments of approximately equal length excluding punctuation and dialogue turn boundaries. This straightforward segmentation is employed in this paper only for comparing results obtained with a more formal model, called **SEG-UA** for User-Agent Segmentation. This model is based on the dialogue internal structure that separately considers features of customer (**i**), agent (**j**) and the whole dialogue (**k**).

4. EXPERIMENTAL PROTOCOL

4.1. Spoken Dialogue dataset

The spoken dialogues corpus is a set of automatically transcribed human-human telephone conversations from the customer care service (CCS) of the RATP Paris transportation system. This corpus comes from the DECODA project [6] and is employed to evaluate the effectiveness of the proposed QMLP on a conversation theme identification task. The DECODA corpus is composed of 1,242 telephone conversations, which corresponds to about 74 hours of signal. The data set was split in 8 categories or themes as described in Table 1.

Table 1. DECODA dataset.

Class label	Number of samples		
	training	development	testing
problems of itinerary	145	44	67
lost and found	143	33	63
time schedules	47	7	18
transportation cards	106	24	47
state of the traffic	202	45	90
fares	19	9	11
infractions	47	4	18
special offers	31	9	13
Total	740	175	327

The LIA-Speeral Automatic Speech Recognition (ASR) system [19] is used to extract textual content of dialogues from the DECODA corpus. Acoustic model parameters were estimated from 150 hours of speech in telephone conditions. The vocabulary contains 5,782 words. A 3-gram language model (LM) was obtained by adapting a basic LM with the training set transcriptions. This system reaches an overall Word Error Rate (WER) of 45.8% on the training set, 59.3% on the development set, and 58.0% on the test set. A “stop list” of 126 words¹ was used to remove unnecessary words (mainly function words) which results in a Word Error Rate (WER) of 33.8% on the training, 45.2% on the development, and 49.5% on the test. These high WER are mainly due to speech disfluencies and to adverse acoustic environments (for example, calls from noisy streets with mobile phones).

4.2. LDA topic spaces and MLP configurations

Different configurations for both LDA topic spaces to build input vectors of real (whole document) and quaternion (depending on the document segmentation process) numbers are performed.

LDA topic spaces configurations

Different LDA models have been learned by varying the number of topics T from 5 to 100. The LDA models also require to choose the hyper-parameters α and β that control the topic distribution in the document and the word distribution in the topics respectively. The standard heuristic is $\alpha = \frac{50}{T}$ and $\beta = 0.01$ [12].

MLP and QMLP configurations

The experiments compare 3 Neural Networks with real (MLP, MLP4) and complex-valued numbers (QMLP). Each neural network has 8 hidden nodes in a single hidden layer :

1. Classical **MLP** with a set of T features (T = number of topics in the LDA space) as input and 8 outputs corresponding to the number of themes contained in the DECODA corpus.
2. Our **QMLP** architecture with inputs from vectors of T quaternions composed of the prior of each topic in the user ($x\mathbf{i}$), agent ($y\mathbf{j}$) and whole document ($z\mathbf{k}$) part respectively to compose the imaginary part of the quaternion (such as image processing task with $R(x\mathbf{i})G(y\mathbf{j})B(z\mathbf{k})$ color code) while the real part r is set to 0. During training, all the coefficients of the output quaternion corresponding to the annotated theme are set to 1 and the coefficients of all the other quaternions are set to zero.
3. For fair comparison, a MLP with the **ijk** parts (the xy and z numbers presented in Equation (14)) of the QMLP concatenated called **MLP4** to compose an input vector for a MLP of the same size of the quaternion is

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

build to evaluate the impact of the Hamilton product during the learning process.

5. RESULTS

The reported results are from a k -fold ($k = 10$) for a robust and convincing comparison between the different document segmentation pre-processes as well as different MLPs (real and quaternion values). Figure 1 reports the accuracies obtained during the theme identification task of spoken dialogues with the MLP, MLP4 and QMLP models from SEG_1 (a) and (b), and SEG-UA (c) and (d) respectively. Table 3 sums up accuracies obtained with the SEG_1 document segmentation while Table 4 presents results with the proposed well-tailored to dialogue SEG-UA document segmentation.

5.1. Document segmentation pre-processes

Two different document segmentations to build input vectors for QMLP are proposed: 1) **SEG_1** is a straightforward segmentation based on the number of words contained in the dialogue with no consideration of the dialogue structure has been proposed in [1]. This representation is based on the topic prior in the four parts of the document (each part contains the same number of words); 2) **SEG-UA** is a proposed more appropriate segmentation in regards to the dialogue internal structure. The quaternion representing a topic of the LDA model, contains the prior of this topic for the user (xi), the agent (yj), and for the whole document to catch the general theme of this document (zk).

Model	Segmentation	Mean Test	Improve. with SEG_1
[1]	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

Table 2. Accuracies in % during the theme identification task with the SEG_1 and SEG-UA document segmentation processes with the first spoken features of Quaternion [1].

The first remark is that the results obtained and reported in Table 2 with the proposed document segmentation well-tailored to dialogue structure (SEG-UA) outperform those from the straightforward segmentation (SEG_1) process regardless of the Neural Network employed with a gain of about 2 and 3 points (last column in Table 2) for MLP4 and QMLP respectively. The difference observed for the MLP for these two document representations is due to the MLP initialization process and the 10-fold employed, but are roughly equivalent (less than 0.6% of difference). Moreover, the gain observed for the QMLP (3.2 points) is greater than the one

obtained with the MLP4 (2.4 points). This is mainly due to the fact that the imaginary coefficients of the same quaternion encode feature dependencies of SEG-UA that are not captured by MLP4. It is also worth emphasizing that the accuracies reached by the MLP and QMLP models are more robust with the SEG-UA segmentation as reported in Figure 1. Indeed, accuracies shown by the curves (a) and (b) for the SEG_1 documents segmentation decrease with LDA models containing a number of topics T greater than 70. This does not appear with the proposed SEG-UA (curves (c) and (d) in Figure 1). We also separately evaluated topic-based representations of users and agents, and obtained a maximum accuracy for the test of 66% and 73.5% for the customer and agent turns respectively. Furthermore, with SEG-UA data and the number of topics between 40 and 45, very high accuracies are observed with the DEV corpus as well as with the TEST corpus. This makes it possible to use the DEV results for predicting the number of topics with which the maximum accuracy is reached with the TEST. These results confirm that separate features for the agent are effective to characterize the fact that the agent task is to focus conversations to a specific theme.

5.2. QMLP vs. MLP

It is worth underlying first that QMLP obtains better performances with the SEG-UA than the MLP and MLP4 with a gain of 2 and 1 points respectively for the Mean Test between all topic models (T from 5 to 100) as shown in Table 4. This remark is also checked for the best accuracy on the development set as well as for the maximum accuracy for the test set.

Model	Max Dev.	Max Test	Mean Test	Epochs
[1]	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

Table 3. Accuracies in % during the theme identification task with the SEG_1 document segmentation and with the first spoken features of Quaternion [1].

The initial intuition behind this study, that the QMLP will obtain better results since the features extracted from the document are suitable for the quaternion representation, is verified. Indeed, Table 3 and Figure 1 (a)-(b) show that QMLP hardly reaches equivalent accuracies of the MLP and MLP4, with a non-relevant document segmentation that reveals little in way of features dependencies in the document. One can easily point out that the number of iterations is smaller (by far) during the learning process of the QMLP, compared to MLP: 7 times quicker for the SEG-UA and 12 times for SEG_1. Although the processing time is a little bit longer with SEG-UA for QMLP, this document segmentation enables this latter to yield better performances.

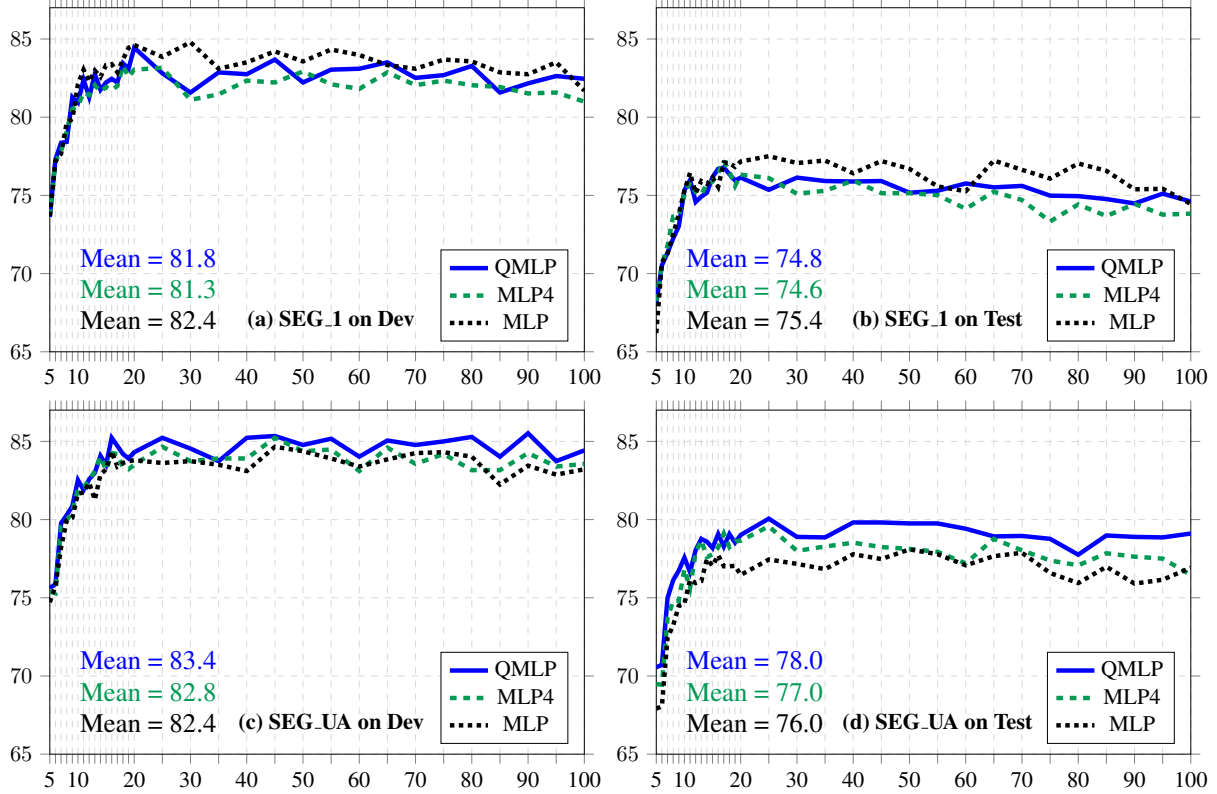


Fig. 1. Accuracies in % obtained by varying the number of topics T in the LDA models for the MLP, MLP4 and QMLP models with the SEG_1 (a)-(b) and SEG_UA (c)-(d) document segmentation processes.

Model	Max Dev.	Max Test	Mean Test	Epochs
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

Table 4. Accuracies in % during the theme identification task with the SEG_UA document segmentation.

Overall, the proposed QMLP catches internal dependencies since the document internal structure is well-coded to build input vectors of quaternions. This is due to the fact that the QMLP algorithm reveals the interpolation between quaternion, and thus, between topics in the document.

6. CONCLUSION

We proposed an original framework based on Quaternion algebra and Neural Networks for real-life document processing. This representation takes advantage of well-segmented documents while MLP fails to extract internal structure based on features dependencies. Both real-valued and quaternion-valued multi-layer perceptrons are compared during a common theme identification task of spoken dialogues. The

results obtained demonstrate that our quaternions MLP (QMLP) reduces the number of epochs needed during the learning process alongside to give a better accuracy than the classical MLP. Moreover, the initial intuitions that: 1) document segmentation is crucial to build robust features and 2) Hamilton product allows the QMLP to reveal document topic statistical dependencies have been demonstrated.

The future work will consist in firstly propose a GPU implementation for QMLP in order to minimize both processing time (CPU) and computation time (GPU). Then, Deep Quaternion Neural Networks will be investigated with the purpose of truly expose the full potential of both Hamilton product and deep neural structures. Then, as shown above, document segmentation is a critical issue while using quaternion-based models. Thus, we will investigate different relevant document representations based on topics and words to reveal latent information and relations. Furthermore, it will be interesting to compare our QMLP to convolutional and recurrent NN respectively described in [20] and [21] which have recently been very successful to capture such latent observations.

7. REFERENCES

- [1] Mohamed Morchid, Georges Linarès, Marc El-Beze, and Renato De Mori, "Theme identification in telephone service conversations using quaternions of speech features," in *Interspeech*. ISCA, 2013.
- [2] Paolo Arena, Luigi Fortuna, Giovanni Muscato, and Maria Gabriella Xibilia, "Multilayer perceptrons to approximate quaternion valued functions," *Neural Networks*, vol. 10, no. 2, pp. 335–342, 1997.
- [3] Geoffrey Hinton, "A practical guide to training restricted boltzmann machines," *Momentum*, vol. 9, no. 1, pp. 926, 2010.
- [4] S.W.R. Hamilton, *Elements of quaternions*, Longmans, Green, & co., 1866.
- [5] Teijiro Isokawa, Nobuyuki Matsui, and Haruhiko Nishimura, "Quaternionic neural networks: Fundamental properties and applications," *Complex-Valued Neural Networks: Utilizing High-Dimensional Parameters*, pp. 411–439, 2009.
- [6] Frederic Bechet, Benjamin Maza, Nicolas Bigouroux, Thierry Bazillon, Marc El-Beze, Renato De Mori, and Eric Arbillot, "Decoda: a call-centre human-human spoken conversation corpus.," in *LREC*, 2012, pp. 1343–1347.
- [7] Nicholas A Aspragathos and John K Dimitros, "A comparative study of three methods for robot kinematics," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 28, no. 2, pp. 135–145, 1998.
- [8] David M Blei, Andrew Y Ng, and Michael I Jordan, "Latent dirichlet allocation," *the Journal of machine Learning research*, vol. 3, pp. 993–1022, 2003.
- [9] Jerome R. Bellegarda, "Exploiting latent semantic information in statistical language modeling," *Proceedings of the IEEE*, vol. 88, no. 8, pp. 1279–1296, 2000.
- [10] Yoshimi Suzuki, Fumiyo Fukumoto, and Yoshihiro Sekiguchi, "Keyword extraction using term-domain interdependence for dictation of radio news," in *17th international conference on Computational linguistics*. ACL, 1998, vol. 2, pp. 1272–1276.
- [11] Thomas Minka and John Lafferty, "Expectation-propagation for the generative aspect model," in *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 2002, pp. 352–359.
- [12] Thomas L. Griffiths and Mark Steyvers, "Finding scientific topics," *Proceedings of the National academy of Sciences of the United States of America*, vol. 101, no. Suppl 1, pp. 5228–5235, 2004.
- [13] Gregor Heinrich, "Parameter estimation for text analysis," Web: <http://www.arbylon.net/publications/text-est.pdf>, 2005.
- [14] Włodzisław Duch and Norbert Jankowski, "Survey of neural transfer functions," *Neural Computing Surveys*, vol. 2, no. 1, pp. 163–212, 1999.
- [15] I.L. Kantor, A.S. Solodovnikov, and A. Shenitzer, *Hypercomplex numbers: an elementary introduction to algebras*, Springer-Verlag, 1989.
- [16] Jack B Kuipers, *Quaternions and rotation sequences*, Princeton university press Princeton, NJ, USA:, 1999.
- [17] Fuzhen Zhang, "Quaternions and matrices of quaternions," *Linear algebra and its applications*, vol. 251, pp. 21–57, 1997.
- [18] J.P. Ward, *Quaternions and Cayley numbers: Algebra and applications*, vol. 403, Springer, 1997.
- [19] Georges Linares, Pascal Nocéra, Dominique Massonnie, and Driss Matrouf, "The lia speech recognition system: from 10xrt to 1xrt," in *Text, Speech and Dialogue*. Springer, 2007, pp. 302–308.
- [20] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.
- [21] Haşim Sak, Andrew Senior, Kanishka Rao, and Françoise Beaufays, "Fast and accurate recurrent neural network acoustic models for speech recognition," *arXiv preprint arXiv:1507.06947*, 2015.