Quaternion Denoising Encoder-Decoder for Theme Identification of Telephone Conversations

Titouan Parcollet, Mohamed Morchid, Georges Linarès

INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Quaternion Denoising Encoder-Decoder for Theme Identification of Telephone Conversations

Titouan Parcollet, Mohamed Morchid, Georges Linarès

LIA, University of Avignon (France)

{firstname.lastname}@univ-avignon.fr

Abstract

In the last decades, encoder-decoders or autoencoders (AE) have received a great interest from researchers due to their capability to construct robust representations of documents in a low dimensional subspace. Nonetheless, autoencoders reveal little in way of spoken document internal structure by only considering words or topics contained in the document as an isolate basic element, and tend to overfit with small corpus of documents. Therefore, Quaternion Multi-layer Perceptrons (QMLP) have been introduced to capture such internal latent dependencies, whereas denoising autoencoders (DAE) are composed with different stochastic noises to better process small set of documents. This paper presents a novel autoencoder based on both hitherto-proposed DAE (to manage small corpus) and the QMLP (to consider internal latent structures) called "Quaternion denoising encoder-decoder" (QDAE). Moreover, the paper defines an original angular Gaussian noise adapted to the specificity of hyper-complex algebra. The experiments, conduced on a theme identification task of spoken dialogues from the DECODA framework, show that the QDAE obtains the promising gains of 3% and 1.5% compared to the standard real valued denoising autoencoder and the QMLP respectively.

Index Terms: Spoken language understanding, Neural networks, Quaternion algebra, Denoising encoder-decoder neural networks

1. Introduction

A basic encoder-decoder neural network [1] (AE) consists of two neural networks (NN): an encoder that maps an input vector into a low-dimensional and fixed context vector; a decoder that generates a target vector by reconstructing this context vector. Multidimensional data such as latent structures of spoken dialogue are difficult to capture by traditional autoencoders due to the unidimensionality of real numbers employed. [2], [3] have introduced a quaternions-based multilayer perceptron (QMLP) as well as a specific spoken dialogues segmentation to better capture internal structures as a result of the Hamilton dot product [4], and thus achieve better accuracies than real-valued multilayer perceptrons (MLP), on a theme identification task of spoken dialogues. A quaternion encoder-decoder has then been proposed by [5] to take advantage of the multidimensionality of hyper-complex numbers to code existing latents relations between pixel colors. However, both quaternions and real numbers based autoencoders suffer from overfitting and degraded generalization capabilities when dealing with small corpus of documents [6]. Indeed, autoencoders try to map the initial vector in a low-dimensional subspace and are thus highly correlated with the number of patterns to learn. To overcome this drawback, a stochastic encoder-decoder called denoising autoencoders (DAE) have been proposed by [6] and investigated in [7, 8, 9]. Intuitively, a denoising auto-encoder encodes artificially corrupted inputs, and try to reconstruct the initial vector. By learning this noisy representation, DAE tends to better abstract patterns in a reduced robust subspace.

The paper proposes a novel quaternion denoising encoder-decoder (QDAE) that takes into account the internal document structure (such as the QMLP) and is able to manage small corpus (as DAE). Nonetheless, traditional noises, such as additive isotropic Gaussian noise [10] , are elaborated for real-numbers autoencoders. Therefore, we also propose a Gaussian angular noise (GAN) adapted to the quaternion algebra. The experiments on the DECODA telephone conversations framework show the impact of the different noises, alongside to underline the performance of the proposed QDAE over DAE, AE, MLP and QMLP.

The rest of the paper is organized as follows: Section 2 presents the quaternion encoder-decoder and Section 3 details the experimental protocol. The results are discussed in Section 4 before concluding on Section 5.

2. Quaternion Denoising Encoder-Decoder

The proposed QDAE is a denoising autoencoder with quaternion numbers. Section 2.1 details the quaternion properties required for the QAE, and QDAE algorithms are presented in Section 2.2.

2.1. Quaternion algebra

Quaternion algebra $\mathbb{Q}$ is an extension of complex numbers 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. A quaternion Q is written as $Q = r1 + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$. In a quaternion, r is the real part while $x\mathbf{i}+y\mathbf{j}+z\mathbf{k}$ is the imaginary part (I) or the vector part. A set of basic quaternion properties needed for the QDAE definition are defined as follow:

- all products of $\mathbf{i},\mathbf{j},\mathbf{k}$: $\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = \mathbf{ijk} = -1$
- quaternion conjugate Q^* of Q is: $Q^* = r1 - x\mathbf{i} - y\mathbf{j} - z\mathbf{k}$
- inner product between two quaternions Q_1 and Q_2 is $\langle Q_1, Q_2 \rangle = r_1r_2 + x_1x_2 + y_1y_2 + u_1u_2$
- normalized of a quaternion $Q^{\triangleleft} = \frac{Q}{\sqrt{r^2+x^2+y^2+z^2}}$
- rotation through the angle of quaternion $R^{\triangleleft}$: $Q' = R^{\triangleleft} Q R^{\triangleleft *}$
- Hamilton product $\otimes$ between Q_1 and Q_2 encodes latent dependencies and is defined as follows:

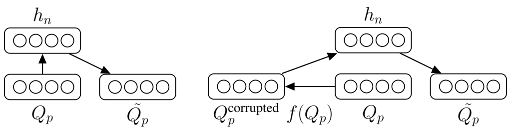
$$\begin{aligned}
Q_1 \otimes Q_2 =& (r_1r_2 - x_1x_2 - y_1y_2 - z_1z_2)+ \\
& (r_1x_2 + x_1r_2 + y_1z_2 - z_1y_2)\mathbf{i}+ \\
& (r_1y_2 - x_1z_2 + y_1r_2 + z_1x_2)\mathbf{j}+ \\
& (r_1z_2 + x_1y_2 - y_1x_2 + z_1r_2)\mathbf{k}
\end{aligned}$$

http://dx.doi.org/10.21437/Interspeech.2017-1029

$Q_1 \otimes Q_2$ performs an interpolation between two rotations following a geodesic over a sphere in the $\mathbb{R}^3$ space. More about hyper-complex numbers can be found in [4, 11, 12] and about quaternion algebra in [13].

2.2. Quaternion Autoencoder (QAE)

The QAE is a three-layered neural network made of an encoder and a decoder (see Figure 1-(a)). The well known autoencoder (AE) is obtained with the same algorithm but with real numbers.



(a) Quaternion autoencoder (b) Quaternion denoising autoencoder

Figure 1: *Illustration of the Quaternion autoencoders.*

Given a set of P normalized inputs $Q_p^{\triangleleft}$ (referenced as Q_p for convenience) ($1 \leq p \leq P$) of size M, the **encoder** computes an hidden representation h_n of $Q_p = \{Q_m\}_{m=1}^{M}$ (N is the number of hidden units):

$$h_n = \alpha(\sum_{m=1}^{M} w_{nm}^{(1)} \otimes Q_m + \theta_n^{(1)})$$

where $w^{(1)}$ is a $N \times M$ weight matrix and $\theta^{(1)}$ is a N-dimensional bias vector; $\alpha(Q)$ is the sigmoid activation function of the quaternion Q [14] $\alpha(Q) = sig(r)1 + sig(x)\mathbf{i} + sig(y)\mathbf{j} + sig(z)\mathbf{k}$, with

$$sig(.) = \frac{1}{1+e^{-.}}. \tag{1}$$

The **decoder** attempts to reconstruct the input vector Q_p from the hidden vector h_n to obtain the output vector $\tilde{Q}_p = \left\{\tilde{Q}_m\right\}_{m=1}^{M}$:

$$\tilde{Q}_m = \alpha(\sum_{n=1}^{N} w_{mn}^{(2)} \otimes h_n + \theta_m^{(2)})$$

where the reconstructed quaternion vector $\tilde{Q}_p$ is M- dimensional, $w^{(2)}$ is a $M \times N$ weight matrix and $\theta^{(2)}$ is a M-dimensional bias vector. During learning, the QAE attempts to reduce the reconstruction error e between $\tilde{Q}_p$ and Q_p by using the traditional Mean Square Error (MSE) [15]:

$$e_{\text{MSE}}(\tilde{Q}_m, Q_m) = ||\tilde{Q}_m - Q_m||^2 \tag{2}$$

for minimizing the total reconstruction error L_{MSE}:

$$L_{\text{MSE}} = \frac{1}{P}\sum_{p\in P}\sum_{m\in M} e_{\text{MSE}}(\tilde{Q}_m, Q_m \tag{3}$$

with respect to the parameters (quaternions) set $\Gamma = \{w^{(1)}, \theta^{(1)}, w^{(2)}, \theta^{(2)}\}$.

2.3. Quaternion Denoising Autoencoder (QDAE)

Traditional autoencoders fail to: 1) separate robust features and relevant information to residual noise [9] from small corpus; 2) take into account the temporal and internal structures of spoken documents. Therefore, denoising autoencoders (DAE) [9] corrupt inputs using specific noises during the encoding and decode this representation to reconstruct the non-corrupted inputs. DAE models learn a robust generative model to better represent small sized corpus of documents; [2] propose to learn internal and temporal structure representation with a quaternion multilayer perceptron (QMLP). The paper proposes to address issues related to small sized corpus (such as DAE) and to temporal structure (QMLP) by introducing a quaternion denoising autoencoder called QDAE. Figure 1-(b) shows an input vector Q_p artificially corrupted by a noise function $f()$ applied to each index Q_m of Q_p as:

$$f(Q_p) = \{f(Q_1), \ldots, f(Q_m), \ldots, f(Q_M)\}. \tag{4}$$

Standard real-numbers-adapted noises :

- *Additive isotropic Gaussian (G)*: Adds a different Gaussian noise to each input values $(Q_1, \ldots, Q_m, \ldots, Q_M)$ of a fixed proportion of patterns Q_p with means and variances of the Gaussian distribution bounded by the corresponding average of all the patterns of the same prediction theme of Q_p.
- *Salt-and-pepper (SP)*: fixes amount of patterns of all patterns Q_p randomly set to 1 or 0.
- *Dropout(D)*: fixes amount of patterns of all patterns Q_p randomly set to 0.

Given a noise function $f()$ the corresponding corrupted quaternion of $Q_m = r1 + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ is:

$$Q_m^{\text{corrupted}} = f(Q_m) = f(r)1 + f(x)\mathbf{i} + f(y)\mathbf{j} + f(z)\mathbf{k}. \tag{5}$$

Nonetheless, such a representation does not take into account the specificity of quaternion algebra since they were designed for real numbers. Indeed, a quaternion represents a rotation over the $\mathbb{R}^3$ space. Therefore, basic additive and non-angular noises such as a Gaussian noise, only represents a one dimensional translation and does not take advantage of the rotation defined by a quaternion.

Quaternion Gaussian Angular Noise (GAN):
The GAN takes advantage of the quaternion algebra (rotation) and is proposed to address the drawback of a weakly adapted noise function (add a noise to each quaternion) to the rotation definition of quaternions. The GAN noise function is based on the rotation of a quaternion vector Q_p around an axis defined in a cone centered in m_t and delimited by v_t; where m_t is the mean and v_t is the variance of the patterns Q_p belonging to theme t. Let R_p^t be a Gaussian noised Quaternion for the theme t defined as:

$$R_p^t = m_t + \mathcal{N}(0, I)\, v_t\ . \tag{6}$$

The Gaussian angular noise function $f()$ rotates Q_p belonging to the theme t around R_p^t to obtain the corrupted Quaternion $Q_p^{\text{corrupted}}$:

$$f(Q_p) = \frac{R_p^t \otimes Q_p \otimes R_p^{t*}}{|R_p^t \otimes Q_p|} \text{ and} \tag{7}$$

$$f(Q_p) = \begin{cases} Q_p, & \text{if } R_p^t = Q_p \\ Q_p^{\text{corrupted}}, & \text{otherwise} \end{cases} \tag{8}$$

It is worth noticing in eq.(8) that f is idempotent since $R_p^t = Q_p$ to maintain the dialogue pattern unaltered.

3. Experimental protocol

The effectiveness of the proposed QDAE-GAN is evaluated during a theme identification task of telephone conversations from the DECODA corpus detailed in Section 3.1. Section 3.2 expresses the dialogue features employed as inputs of autoencoders as well as the configurations of each neural network.

3.1. Spoken Dialogue dataset

The DECODA corpus [16] contains human-human telephone real-life conversations collected in the CSS of the Paris transportation system (RATP). It is composed of $1,242$ telephone conversations, corresponding to about 74 hours of signal, split into a train (740 dialogues), a development (dev - 175 dialogues) and a test set (327 dialogues). Each conversations is annotated with one of 8 themes. Themes correspond to customer problems or inquiries about itinerary, lost and found, time schedules, transportation cards, state of the traffic, fares, fines and special offers. The LIA-Speeral Automatic Speech Recognition (ASR) system [17] is used for automatically transcribing each conversation. Acoustic model parameters are estimated from 150 hours of telephone speech. The vocabulary contains $5,782$ words. A 3-gram language model (LM) is obtained by adapting a basic LM with the training set transcriptions. Automatic transcriptins are obtained with word error rates (WERs) of 33.8%, 45.2% and $49.\%$ on the train, dev. and test sets respectively. These high rates are mainly due to speech disfluencies in casual users and to adverse acoustic environments in metro stations and streets.

3.2. Input features and Neural Networks settings

The experiments compare our proposed QDAE with DAE based on real-numbers [7] and to the QMLP[2].

Input features: [2] show that a LDA [18] space with 25 topics and a specific user-agent document segmentation involving the quaternion $Q = r1 + x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ to be build with the user part of the dialogue in the first complex value x, the agent in y and the topic prior of the whole dialogue on z, achieve the best results on 10 folds with the QMLP. Therefore, we keep this segmentation and concatenate the 10 representations of size 25 in a single input vector of size $M = 250$. Indeed, the compression of 10 folds in a single input vector gives to DAEs more features for generalizing patterns. For fair comparison, a QMLP with the same input vector is tested.

QDAE and QMLP configurations: The appropriate size of the hidden layer h for the QDAE have to be chosen by varying the number of neurons of the hidden layer to change the amount and the shape of features given to the classifier. Different autoencoders have thus been learned by fluctuating the hidden layer size from 10 to 120. Finally a QMLP classifier is trained with 8 hidden neurons; the hidden layer of the QAE, QDAE as the input vectors; and 8 outputs neurons (8 themes t on the DECODA corpus).

4. Experiments and Results

The proposed Quaternion denoising autoencoder (QDAE) is compared to the quaternion autoencoder (QAE) in Section 4.1, throughout a theme identification task of telephone conversations described in Section 3.1. For fair comparison, the QDAE is then compared to the real-valued AE and MLP in Section 4.2.

4.1. QDAE with additive and angular noises

Figure 2 shows the accuracies obtained with the denoising quaternion encoder-decoder for the development and the test set during the theme identification task of telephone conversations of DECODA project.

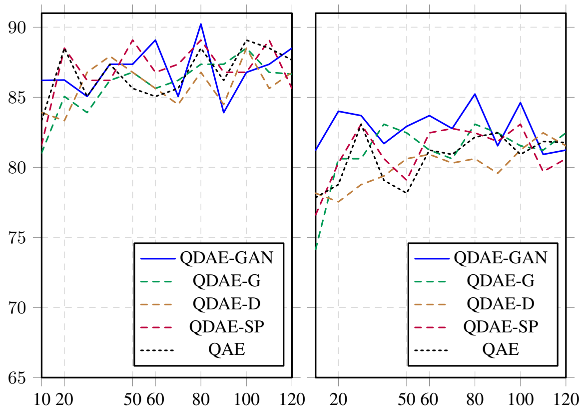


Figure 2: ***Accuracies*** *in $\%$ obtained on the development (left) and test (right) set by varying the number of neurons in the hidden layer of the QAE and QDAE.*

The first remark is that the results obtained on the development dataset reported in Fig.2 are similar whatever the model employed. Nonetheless, the proposed QDAE-GAN gives better, and more robust to hidden layer size variation, results on the test dataset than any other methods.

Table 1: **Accuracies** *in % obtained by proposed quaternion encoder-decoders on the DECODA dataset*

Models	Dev.	Best Test	Real Test
QAE	89.1	83.0	80.9
QDAE-SP	88.5	82.5	81.2
QDAE-G	88.5	83.1	81.5
QDAE-D	89.1	83.0	82.5
QDAE-GAN	**90.2**	**85.2**	**85.2**

Table 1 validates the results observed for the QDAE-GAN with a gain of more than 3.5% and 2.5% for QDAE-G and QDAE-D respectively. As expected, traditional noises give worse results compared to the adapted noise due to the specificities of the quaternion algebra. Indeed, an additive real-based Gaussian noise applied to a quaternion does not take advantage of rotations defining quaternions. It is worth underlying the bad performances reported with the QDAE-SP and QDAE-D, which are not based on real or quaternion algebra specificities: These poor performances are explained by the high impact of zero values propagated during the Hamilton product (see Section 2.1) by increasing the number of dead neurons through the neural network. Finally, the non-corrupted QAE gives a good "best test" value on the test dataset (83%) regarding the other QDAE, proving the non-relevance of real-based noises to quaternion-based autoencoders.

4.2. QDAE *vs.* real-valued neural networks

For a fair comparison this original QDAE-GAN approach is compared to real-valued autoencoders and traditional neural networks, and the results are depicted in Table 2.

Table 2: **Summary** *of accuracies in % obtained by different neural networks on the DECODA famework.*

Models	Type	Dev.	Best Test	Real Test	Impr.
MLP[2]	$\mathbb{R}$	85.2	79.6	79.6	-
QMLP	$\mathbb{Q}$	89.7	83.7	83.7	+4.1
AE[7]	$\mathbb{R}$	-	-	81	-
QAE	$\mathbb{Q}$	89.1	83.0	80.9	-0.1
DAE[7]	$\mathbb{R}$	-	-	74.3	-
DSAE[7]	$\mathbb{R}$	88.0	83.0	82.0	+7.7
QDAE-GAN	$\mathbb{Q}$	**90.2**	**85.2**	**85.2**	**+10.9**

Table 2 shows that non-adapted noise and standard QAE give worse performances than a QMLP because of the lack of unseen compressed information they give to the classifier. It is worth emphasizing that the best accuracies observed are obtained by the QDAE-GAN representing a gain of 11% regarding DAE [7]. The results depicted on Table 2 demonstrate the global improvement of performances of the quaternion-valued neural networks compared to the real-valued ones. Indeed, QMLP also gives a important gain of more than 4% regarding the MLP; QDAE-GAN obtains a gain of 3.2% compared to DSAE.

5. Conclusion

Summary. This paper proposes a promising denoising encoder-decoder based on the quaternion algebra coupled with an original and well-adapted quaternion Gaussian angular noise. The initial intuition that the QDAE better captures latent relations between input features and can generalize from small corpus, has been demonstrated. It has been shown that ongoing noises during learning must be adapted to the quaternion algebra to give better results and truly expose the full potential of quaternion neural networks. Moreover, this paper shows that quaternion-valued neural networks always perform better than real-valued ones achieving impressive accuracies on the small DECODA corpus with less input features and less neural parameters.

Limitations and Future Work. Document segmentation is a crucial issue when it comes to better capture latent, temporal and spacial information and thus needs more investigation to expose the potential of quaternion-based models. Moreover, the lack of GPU tools to manage quaternions impline a massive implementation time to deal with bigger spoken document corpus. A future work is to investigate other quaternion adapted noises, and other quaternion based neural networks which better take into consideration the document internal structure , such as recurrent neural networks and Long Short Term Memory neural networks.

6. References

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[2] T. Parcollet, M. Morchid, P.-M. Bousquet, R. Dufour, G. Linarès, and R. De Mori, "Quaternion neural networks for spoken language understanding," in *Spoken Language Technology Workshop (SLT), 2016 IEEE*. IEEE, 2016, pp. 362–368.

[3] M. Morchid, G. Linarès, M. El-Beze, and R. De Mori, "Theme identification in telephone service conversations using quaternions of speech features," in *Interspeech*. ISCA, 2013.

[4] I. Kantor, A. Solodovnikov, and A. Shenitzer, *Hypercomplex numbers: an elementary introduction to algebras*. Springer-Verlag, 1989.

[5] T. Isokawa, N. Matsui, and H. Nishimura, "Quaternionic neural networks: Fundamental properties and applications," *Complex-Valued Neural Networks: Utilizing High-Dimensional Parameters*, pp. 411–439, 2009.

[6] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 1096–1103.

[7] K. Janod, M. Morchid, R. Dufour, G. Linares, and R. De Mori, "Deep stacked autoencoders for spoken language understanding," *Matrix*, vol. 1, p. 2, 2016.

[8] X. Lu, Y. Tsao, S. Matsuda, and C. Hori, "Speech enhancement based on deep denoising autoencoder." in *Interspeech*, 2013, pp. 436–440.

[9] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *Journal of Machine Learning Research*, vol. 11, no. Dec, pp. 3371–3408, 2010.

[10] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[11] J. B. Kuipers, *Quaternions and rotation sequences*. Princeton university press Princeton, NJ, USA:, 1999.

[12] F. Zhang, "Quaternions and matrices of quaternions," *Linear algebra and its applications*, vol. 251, pp. 21–57, 1997.

[13] J. Ward, *Quaternions and Cayley numbers: Algebra and applications*. Springer, 1997, vol. 403.

[14] P. Arena, L. Fortuna, G. Muscato, and M. G. Xibilia, "Multilayer perceptrons to approximate quaternion valued functions," *Neural Networks*, vol. 10, no. 2, pp. 335–342, 1997.

[15] Y. Bengio, "Learning deep architectures for ai," *Foundations and trends® in Machine Learning*, vol. 2, no. 1, pp. 1–127, 2009.

[16] F. Bechet, B. Maza, N. Bigouroux, T. Bazillon, M. El-Beze, R. De Mori, and E. Arbillot, "Decoda: a call-centre human-human spoken conversation corpus." in *LREC*, 2012, pp. 1343–1347.

[17] G. Linares, P. Nocéra, D. Massonie, and D. Matrouf, "The lia speech recognition system: from 10xrt to 1xrt," in *Text, Speech and Dialogue*. Springer, 2007, pp. 302–308.

[18] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation," *the Journal of machine Learning research*, vol. 3, pp. 993–1022, 2003.