

# Stage de recherche

---

Réseaux de neurones Multi-échelles Croissants

**Esteban Marquer**

*Année 2017–2018*

Projet réalisé pour l'équipe SYNALP du laboratoire LORIA

Maître de stage : Christophe Cerisara

Encadrant universitaire : Jeanine Souquières



# Stage de recherche

---

Réseaux de neurones Multi-échelles Croissants

**Esteban Marquer**

**Année 2017–2018**

Projet réalisé pour l'équipe SYNALP du laboratoire LORIA

Esteban Marquer

[marquer.esteban@etu.univ-lorraine.fr](mailto:marquer.esteban@etu.univ-lorraine.fr)

Institut des Sciences du Digital Management & Cognition

193 avenue Paul Muller,

CS 90172, VILLERS-LÈS-NANCY

+33 (0)3 72 74 16 18

[idmc-contact@univ-lorraine.fr](mailto:idmc-contact@univ-lorraine.fr)

LORIA

Campus scientifique

BP 239

54506, Vandoeuvre-lès-Nancy Cedex

+33 (0)3 83 59 20 00



Encadrant : Christophe Cerisara



## Remerciements

Je tiens à remercier M. Christophe Cerisara, qui a élaboré un sujet passionnant pour ce stage, et qui a su m'accompagner tout au long de cette aventure, malgré un emploi du temps chargé et des responsabilités nombreuses en tant que chef d'équipe.

Je remercie M. Samuel Cruz-Lara, Mme. ??? et Mme. ??? avec qui j'ai brièvement collaboré au sein du projet PAPUD.

Je remercie aussi que Mme. Jeanine Souquières, qui m'a conseillé lors de l'élaboration de ce rapport.

Je tiens tout particulièrement à remercier M. Maxime Amblard, sans qui je n'aurais pas obtenu ce stage.

Mais je remercie aussi tous les stagiaires et doctorants qui ont supporté mon humour pendant plus de 3 mois, ainsi que les stagiaires de l'IDMC qui le supportent depuis bien plus longtemps.

Enfin, je remercie Annick Jacquot, qui s'est chargé de toutes les questions administratives de mon stage ; Caroline et toute l'équipe de la cafétéria, qui m'ont offert un cadre inoubliable ; et tous les gens du LORIA qui m'ont offert un accueil chaleureux et qui font vivre ce laboratoire.



## Avant-propos

La lecture du présent mémoire ne nécessite aucune connaissance préalable en Apprentissage Automatique ou en Apprentissage Profond.

Cependant, de nombreux termes techniques sont utilisés. La plupart est abrégée ou sous forme de sigle.

D'une part, ces abréviations sont habituellement en anglais dans la littérature du domaine. Aussi, pour maintenir la lisibilité pour les lecteurs initiés au domaine, les abréviations utilisées seront sous leur forme anglaise. Pour maintenir une cohérence dans les termes utilisés, les expressions dont les versions anglaises et françaises. Enfin, les termes techniques seront toujours introduits en français, accompagnés d'une explication, de la traduction anglaise et du sigle anglais. Par exemple : Exemple de Terme Technique (*Technical Expression Example* en anglais, TEE).

D'autre part, une partie du rapport est dédiée à l'explication des termes et concepts utilisés, et un glossaire est présent en fin d'ouvrage.





# Table des matières

<b>Remerciements</b>	<b>iii</b>
<b>Avant-propos</b>	<b>v</b>
<b>1 Introduction</b>	<b>3</b>
1.1 Contexte & Enjeux . . . . .	3
1.2 Objectifs . . . . .	3
1.3 Plan . . . . .	4
<b>2 Présentation du laboratoire et de l'équipe</b>	<b>5</b>
2.1 SYNALP . . . . .	5
2.2 LORIA . . . . .	5
2.3 ITEA3/PAPUD . . . . .	5
<b>3 Projet</b>	<b>7</b>
3.1 Termes et concepts . . . . .	8
3.2 short title . . . . .	9
3.3 Autre section . . . . .	9
<b>4 Exemples Listings</b>	<b>11</b>
<b>5 Conclusion</b>	<b>15</b>
<b>Annexes</b>	<b>17</b>



# 1 Introduction

## 1.1 Contexte et enjeux du stage

D'une part, depuis quelques années, les "réseaux de neurones" ont connu une explosion de popularité. Ce qui se cache derrière cette hype est la combinaison de théories relativement anciennes et de avancées technologiques permettant la mise en œuvre desdites théories. Un des intérêts de ces méthodes d'Apprentissage Profond (*Machine Learning* en anglais, ML) est la capacité à apprendre à partir de données restreintes. Un des domaines exploitant les performances de ces outils est le TAL, au vu de ça et de ça. Une application au TAL de l'apprentissage différentiel est le LM qui nous intéressera particulièrement dans ce mémoire.

D'autre part, nous sommes à l'ère du "Big Data". Cela implique que les quantités de données exploitables produites de nos jours est très grande. Dans ce rapport, nous nous intéresserons à des volumes de données bien au-delà des volumes habituellement utilisés dans le domaine.

Ainsi, nous nous explorerons la question de l'application des méthodes du DL du point de vue du TAL sur des ensembles de données de grande taille.

L'axe principal de ce mémoire est l'application de telles méthodes sur des gros volumes de données. Cela implique à la fois des problématiques relativement classiques en NN d'architecture du réseau, et de choix d'algorithme d'entraînement ; mais aussi des questions plus pragmatiques d'optimisation liées au volume de données.

## 1.2 Objectifs du stage

Deux objectifs successifs se distinguent dans le stage.

L'objectif initial du stage est d'explorer une idée d'architecture de Réseau de Neurones Artificiels (*Neural Network* en anglais, NN) innovante, imaginée par notre maître de stage Mr Cerisara. Par la même occasion, ce stage est l'opportunité pour moi d'apprendre à manipuler les NN.

À l'issue du deuxième mois du stage, au vu des résultats de l'architecture et de l'évolution du contexte, la mission du stage a aussi évolué.

Le nouvel objectif est la réalisation d'un NN et des outils nécessaires à son utilisation, en mettant à profit les connaissances acquises durant la première partie du stage. Cette réalisation servira de base technique pour une partie du projet PAPUD (???).

Les tenants et aboutissants des deux objectifs seront expliqués en détails dans le chapitre 3.

## 1.3 Plan du rapport

Dans un premier temps, nous avons présenté à la fois le contexte, les enjeux, et les objectifs généraux du stage. Dans un second temps, nous allons nous attarder plus en détail sur les différentes entités impliquées. Dans un troisième temps, nous allons décrire plus en détail les deux aspects du stage : l'idée d'architecture de NN et l'intérêt d'une telle architecture d'un côté, et le projet PAPUD, ses implications et la portée du stage dans ce projet. Nous verrons aussi en quoi le projet PAPUD est dans la continuité de la première partie du stage. Dans un quatrième temps, nous exposerons le déroulement pas-à-pas du stage, avec les obstacles rencontrés, la façon de les surmonter, et en quoi chaque résultat entraîne l'étape suivante. Dans un cinquième et dernier temps, nous ferons une rétrospective sur l'avancement des objectifs, la qualité des résultats obtenus, et les apports du stage.

## **2 Présentation du laboratoire et de l'équipe**

### **2.1 L'équipe SYNALP**

L'équipe SyNaLP (Symbolic and statistical NLP) est une équipe de recherche affiliée à la fois au Centre National de la Recherche Scientifique (CNRS) et à l'Université de Lorraine (UL).

SYNALP est dans le

Des informations détaillées sont disponible sur le site de l'équipe (en Anglais) [1]

### **2.2 Présentation du LORIA**

Le Laboratoire Lorrain d'Informatique et ses Applications (LORIA) est une Unité Mixte de Recherche (UMR 7503), commune à plusieurs établissements : le CNRS, l'UL et l'Institut National de Recherche en Informatique et en Automatique (INRIA).

Le Loria a pour mission la recherche fondamentale et appliquée en sciences informatiques et ce, depuis sa création, en 1997.

### **2.3 Présentation du projet ITE3-PAPUD, cas d'utilisation BULL**

**Initiative ITEA**

**Projet PAPUD**

**Cas d'utilisation BULL**

Par la suite, nous désignerons ce projet par "Projet PAPUD"



### **3   Projet**

## **3.1 Terminologie et concepts fondamentaux**

### **3.1.1 lm**

**char-lm**

### **3.1.2 nn**

### **3.1.3 rnn**

**lstm**

### **3.1.4 msnn**

**gmsnn**



## 3.2 Du char-lstm au char-gmsnn

see <https://itea3.org/project/papud.html>

## 3.3 Autre section

Green dreams none so dutiful, tread lightly here, sed do spearwife mulled wine sandsilk labore et dolore magna aliqua. Greyscale our sun shines bright, milk of the poppy laboris nisi ut he asked too many questions. Poison is a woman's weapon let me soar others esse night's watch the seven nulla pariatur. Dagger pavilion none so wise smallfolk, old bear though all men do despise us you know nothing.

### 3.3.1 Première sous-section

#### Première sous-sous section

Exemple d'illustration :



FIGURE 3.1 – Logo de L'UL

La Figure 3.1 représente le logo de Institut des Sciences du Digital Management & Cognition. Ceci est une référence bibliographique [GOT4].



## 4 Exemples Listings

Il est aisé d'insérer du code dans un rapport. Il suffit de définir le langage, la légende à afficher et enfin un Label pour pouvoir y faire référence. Le résultat est donnée dans le listing 4.1. Il est également possible de changer les couleurs, pour cela il faut éditer le lstset dans la classe tnreport.cls.

```
1 void CEquation::IniParser()
2 {
3   if (!pP){ //if not already initialized ...
4     pP = new mu::Parser;
5
6     pP->DefineOpert("%", CEquation::Mod, 6); //deprecated
7     pP->DefineFun("mod", &CEquation::Mod, false);
8     pP->DefineOpert("&", AND, 1); //DEPRECATED
9     pP->DefineOpert("and", AND, 1);
10    pP->DefineOpert("|", OR, 1); //DEPRECATED
11    pP->DefineOpert("or", OR, 1);
12    pP->DefineOpert("xor", XOR, 1);
13    pP->DefineInfixOpert("!", NOT);
14    pP->DefineFun("floor", &CEquation::Floor, false);
15    pP->DefineFun("ceil", &CEquation::Ceil, false);
16    pP->DefineFun("abs", &CEquation::Abs, false);
17    pP->DefineFun("rand", &CEquation::Rand, false);
18    pP->DefineFun("tex", &CEquation::Tex, false);
19
20    pP->DefineVar("x", &XVar);
21    pP->DefineVar("y", &YVar);
22    pP->DefineVar("z", &ZVar);
23  }
24 }
```

Fragment de code 4.1 – Premier Exemple

Il est également possible d'afficher du code directement depuis un fichier source, le résultat de cette opération est visible dans le listing ??

De nombreux langages sont supportés :

ABAP2,4, ACSL, Ada4, Algol4, Ant, Assembler2,4, Awk4, bash, Basic2,4, C#5, C++4, C4, Caml4, Clean, Cobol4, Comal, csh, Delphi, Eiffel, Elan, erlang, Euphoria, Fortran4, GCL, Gnuplot, Haskell, HTML, IDL4, inform, Java4, JVMIS, ksh, Lisp4, Logo, Lua2, make4, Mathematica1,4, Matlab, Mercury, MetaPost, Miranda, Mizar, ML, Modelica3, Modula-2, MuPAD, NASTRAN, Oberon-2, Objective C5 , OCL4, Octave, Oz, Pascal4, Perl, PHP, PL/I, Plasm, POV, Prolog, Promela, Python, R, Reduce, Rexx, RSL, Ruby, S4, SAS, Scilab, sh, SHELXL, Simula4, SQL, tcl4, TeX4, VBScript, Verilog, VHDL4, VRML4, XML, XSLT.

Il est néanmoins possible de définir le sien, il faudra alors ajouter dans la classe `tnreport.cls` du code ressemblant au listing 4.2. On y définit les différents mots-clés, ainsi que les délimiteurs des chaînes de caractère et des commentaires.

```
1 \lstdefinelanguage{amf}
2 {keywords=
3 {
4 xml,
5 amf,
6 volume,
7 material,
8 coordinates,
9 vertices,
10 vertex,
11 triangle,
12 x,
13 y,
14 z,
15 v1,
16 v2,
17 v3,
18 mesh,
19 object,
20 constellation,
21 metadata,
22 color,
23 texmap,
24 texture,
25 utex1,
26 utex2,
27 utex3,
28 instance,
29 deltax,
30 deltay,
31 deltaz,
32 r,
33 g,
34 b,
35 rx,
36 ry,
37 rz,
38 composite
39 },
40 sensitive=false,
41 morestring=[b]",
42 comment=[s]{<!--}{-->}
43 }
```

Fragment de code 4.2 – Syntaxe définition d'un langage



## **5 Conclusion**





# ***Annexes***



# **Table des annexes**

<b>Bibliographie / Webographie</b>	<b>21</b>
<b>Liste des illustrations</b>	<b>23</b>
<b>Liste des tableaux</b>	<b>25</b>
<b>Entités, Instituts et???</b>	<b>29</b>
<b>Acronymes</b>	<b>30</b>
<b>Glossaire</b>	<b>31</b>
<b>A Rapports d'avancement du premier projet</b>	<b>33</b>
<b>B Rapports d'avancement du projet PAPUD</b>	<b>35</b>
<b>C Copie de la convention de stage</b>	<b>37</b>
<b>D Copie de l'avenant à la convention de stage</b>	<b>39</b>



## Bibliographie / Webographie

- [1] *Synalp*. [Online; accessed 30. Jul. 2018]. 2018. URL : <http://synalp.loria.fr/pages/about-synalp> (cf. p. 5).



## Liste des illustrations

3.1	Logo de L'UL . . . . .	9
-----	------------------------	---





## Liste des tableaux



# Table des annexes

<b>Bibliographie / Webographie</b>	<b>21</b>
<b>Liste des illustrations</b>	<b>23</b>
<b>Liste des tableaux</b>	<b>25</b>
<b>Entités, Instituts et???</b>	<b>29</b>
<b>Acronymes</b>	<b>30</b>
<b>Glossaire</b>	<b>31</b>
<b>A Rapports d'avancement du premier projet</b>	<b>33</b>
<b>B Rapports d'avancement du projet PAPUD</b>	<b>35</b>
<b>C Copie de la convention de stage</b>	<b>37</b>
<b>D Copie de l'avenant à la convention de stage</b>	<b>39</b>



## Entités, Instituts et ???

**CNRS** Centre National de la Recherche Scientifique 5

**INRIA** Institut National de Recherche en Informatique et en Automatique 5

**LORIA** Laboratoire Lorrain d'Informatique et ses Applications 5

**PAPUD** 3, 19, 27, 35

**SYNALP** 5

**UL** Université de Lorraine 5

# Acronymes

**Char-GMSNN-LM** *Glossaire* : Char-GMSNN-LM

**Char-LM** *Glossaire* : Char-LM

**DL** *Glossaire* : DL

**GMSNN** *Glossaire* : GMSNN

**GPU** *Glossaire* : GPU

**LM** *Glossaire* : LM

**ML** *Glossaire* : ML, 3

**MSNN** *Glossaire* : MSNN

**NN** *Glossaire* : NN, 3

**RNN** *Glossaire* : RNN

# Glossaire

**Char-GMSNN-LM** Un Modèle du Langage au niveau du Caractère basé sur un Réseau de Neurones Multi-Échelles Croissant. C'est un modèle du langage basé sur un réseau de neurones artificiels. Ce réseau de neurones est un GMSNN utilisé comme Char-LM. *voir* Char-LM & GMSNN

**Char-LM** Un Modèle de la Langue au niveau du Caractère (*Character-level Language Model* en anglais) est un Modèle de la Langue qui prédit non pas le prochain mot à partir des mots précédents, mais le prochain caractère à partir des caractères précédents *voir* LM

**DL** L'Apprentissage Profond (*Deep Learning* en anglais) est une méthode de Machine Learning utilisant la technique des réseaux de neurones. *voir* ML

**GMSNN** (le nombre de couches augmente selon le nombre d'entrées) *voir* MSNN

**GPU** Un processeur graphique (*Graphical Processing Unit* en anglais) est un composant d'ordinateur spécialisé, qui montre d'excellentes performances dans les calculs impliquant des matrices (ex. : images) *voir* matrice

**matrice** 31

**ML** L'Apprentissage Automatique (*Machine Learning* en anglais) est un ensemble de "méthodes [statistiques] permettant à une machine (au sens large) d'évoluer par un processus systématique, et ainsi de remplir des tâches difficiles ou problématiques par des moyens algorithmiques plus classiques". [wikiml] 3

**MSNN** *voir* RNN

**NN** Un Réseau de Neurones Artificiels (*Neural Network* en anglais) est 3

**RNN** Un Réseau de Neurones Artificiels Récurrents, plus simplement Réseau de Neurones Récurrents (*Recurrent Neural Network* en anglais) est un réseau de neurones artificiels suivant une architecture dite récurrente.

Ce genre de réseau est utilisé pour travailler avec des séquences d'entrées et/ou de sorties; il y a transmission d'information entre chaque élément de la séquence. [wikirnn] *voir* NN





## **A Rapports d'avancement du premier projet**



## **B Rapports d'avancement du projet PAPUD**



## **C Copie de la convention de stage**



## **D Copie de l'avenant à la convention de stage**

