MINISTERE DE L’ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE

**I**NSTITUT **N**ATIONAL DES **S**CIENCES **A**PPLIQUEES

TOULOUSE





**Département de Génie Electrique & Informatique**

**PROJET DE FIN D’ETUDES**

TITRE DU PROJET

**Natixis - Paris**

Nom prénom de l’élève

Spécialité – Majeure Juin 2018

# Résumé

Résumé en 15 / 20 lignes du rapport de stage

Dans le cadre de notre programme innovation, à partir de use cases, le collaborateur aura pour mission de trouver des applications du machin learning en particulier réseau de neurones one layer et convolutifs, pour résoudre les problèmes de réconciliation de la banque notamment.

Les sujets de développement concerneront notamment les réconciliations s (1,1), (1/N),(N/M),

# Remerciements

Les personnes à remercier.

# Table des matières

A (re)-générer (Menu Références -> Tables des Matières -> Insérer une Table des Matières)

Doit commencer sur une page recto

[Résumé i](#_Toc409557801)

[Remerciements i](#_Toc409557802)

[Table des matières i](#_Toc409557803)

[Glossaire i](#_Toc409557804)

[Chapitre 1. Introduction 1](#_Toc409557805)

[Chapitre 2. Cadre et objectifs du stage 1](#_Toc409557806)

[Chapitre 3 et suivants. Réalisations 1](#_Toc409557807)

[Chapitre X. Conclusions et perspectives 1](#_Toc409557808)

[Annexes 1](#_Toc409557809)

[Chapitre Y. Consignes sur la rédaction (Titre niveau 1) 1](#_Toc409557810)

[1.1. Partie 1 (Titre niveau 2) 1](#_Toc409557811)

[1.1.1. Sous partie 1 (titre niveau 3) 1](#_Toc409557812)

[1.1.2. Sous partie 2 (titre niveau 3) 1](#_Toc409557813)

[1.2. Partie 2 (Titre niveau 2) 1](#_Toc409557814)

[Bibliographie 1](#_Toc409557815)

[Liste des Illustrations 1](#_Toc409557816)

# Glossaire

Définir les termes importants utilisés dans le rapport

Doit commencer sur une page recto

Réseaux neuronal convolutif

Réseau de neurones récurrents(LSTM)

# Chapitre 1. Introduction

Introduction générale sur le sujet de stage :

**contexte (au sens large) :**

Dans le cadre de notre programme innovation, à partir de use cases, le collaborateur aura pour mission de trouver des applications du machin learning en particulier réseau de neurones one layer et convolutifs, pour résoudre les problèmes de réconciliation de la banque notamment.

Les sujets de développement concerneront notamment les réconciliations s (1,1), (1/N),(N/M),

**société, entités,**

Natixis est une banque de financement, de gestion et de services financiers, créée en 2006, filiale du groupe BPCE.

clients, secteurs d’activités,

problématique générale, gestion de flux financière, flux incoming et outgoing(débit crédit) le system bancaire doit être capable de matcher ces flux équilibrèrent cad somme débit = somme crédit. Mais souvent les flux incoming qui vient d’extérieur ne contient pas beaucoup d’information ou contient des information qu’on n’est pas capable de traiter (par exp les code (trade ID)dans leur system) et le systems traditionel ne sait pas comment les matcher. D’où on a besoin d’intelligence artificiel pour nous aider.

domaines de compétences concernés : analyse financière, programmation,

Présentation générale du rapport (ce que va contenir chaque chapitre). Dans les chapitres suivants je vais presenter l’objectif du stage et le RN qu’on a testé et les analyses

**Pas de présentation de l’entreprise à ce niveau-là (elle sera dans la partie Annexe)**

Environ 2 à 3 pages

# Chapitre 2. Cadre et objectifs du stage

Préciser dans ce chapitre comment se situe votre stage dans la société (faisant d’un projet plus général par exemple, dans ce cas décrire le projet général, faisant partie d’une thématique plus général et dans ce cas expliciter cette thématique, ….).

Complement du system existant. Colaborer avec sys existant. Projet IA

Example flux

Donner les objectifs qui vous ont été fixés (en précisant si vous êtes seul pour réaliser cela ou intégré dans une équipe).

2 stagiaire : objectif .creer un système qui permet de traiter automatiquement les flux suspends.

Identifier les client, matcher les flux, contrôler la cohérence du résultat .

2eme projet : IHM + RN

Il est important de mettre en avant votre travail car c’est cela que l’on cherche à évaluer sans oublier de le situer dans une équipe si c’est le cas.

Environ 10 pages

# Chapitre 3 et suivants. Réalisations

Total : environ 30 pages

A adapter le sens de « réalisations » selon le contexte (cela peut être un état de l’art, un travail de conception, un travail de développement, de validation, de tests, de gestion de projet, ….)

Ne pas forcément tout présenter mais faire en sorte que l’ensemble corresponde à un volume de travail pertinent.

Prise en main : travailler avec l’ancien model en scikt-learn pour comprendre le modèle et les données

Ensuit on trouve que sk learn n’est pas suffisant pour réaliser le RN plus compliqué. Il manque de la liberté de chaque couche par exemple on ne peut pas avoir différente fonction d’activation pour chaque couche. On n’a pas de couche convolutif etc. On trouve un autre outil : Tensorflow, un outil développé par Google et largement utilisé pour le RN

Pour apprendre tensorflow, on essaye de refaire le même modèle avec Tensorflow. On test avec l’ancien jeu de données et vérifier s’il données le même résultat.

Ensuit on essaye de trouver des modèles qui correspondre au notre problématique. (on décompose le problème en 3 parties : reconnaissance du message, matching, contrôle résultat ). On ***pense*** que le problème est un problème de classification du text et on a trouvé le CNN et RNN(surtout LSTM) et la combination de deux CNN+LSTM

On essaye les trois et on trouve que CNN donne la meilleure performance au niveau de la prédiction et aussi dans la phase d’entrainement.

**Model CNN**

Word2vec

CNN sur le NLP

Visualisation du résultat avec tensorflow

Analyse résultat (courbe précision et loss, t-SNE sur word2vec )

Nettoyage du jeu d’entrainement

Retester

**Matcheur**

Règle de matching :

Même dv etc

Règle spéciale pour certaine devise

Analyse résultat

Modifier les règles et ses priorités

**Controleur**

Pour chaque réalisation présentée dans le rapport, expliciter les choix techniques effectués, le travail réalisé, le résultat obtenu, les outils utilisés, …

Bien évidemment, vous n’avez pas forcément la liberté de décider des choix techniques lors de votre PFE, mais dans le rapport on vous demande de prendre du recul sur les choix effectués et de fournir votre analyse de ces choix-là.

Ces chapitres de réalisation ne sont pas une « documentation technique » destinée à la société et n’ont pas vocation à être un « manuel utilisateur ».

Le rapport sert aux enseignants INSA à évaluer votre démarche, la prise de recul ou la synthèse que vous effectuez sur votre travail et l’analyse que vous pouvez mener. Les difficultés rencontrées ou les impasses peuvent être analysées dans ce rapport.

L’approche générale de l'utilisation de CNN pour le traitement du langage naturel. Les phrases sont mappées sur des vecteurs d'incorporation et sont disponibles sous forme de matrice dans le modèle. Les convolutions sont effectuées sur la saisie en utilisant des noyaux de taille différente, par exemple 2 ou 3 mots à la fois. L’ensemble de caractéristique obtenu sont ensuite traitées à l'aide d'une couche de max pooling pour condenser ou résumer les caractéristiques extraites.

Illustration d'une architecture CNN pour la classification des phrases. Nous décrivons trois tailles de région de filtre: 2,3,4, chacune ayant 2 filtres. Les filtres effectuent des convolutions sur la matrice de phrases et génèrent des cartes de caractéristiques (de longueur variable); Le regroupement 1-max est effectué sur chaque carte, c'est-à-dire que le plus grand nombre de chaque carte de caractéristiques est enregistré. Ainsi, un vecteur de caractéristiques univariées est généré à partir des six cartes et ces six entités sont concaténées pour former un vecteur de caractéristiques pour l'avant-dernière couche. Le softmax final reçoit ensuite ce vecteur d'entité en entrée et l'utilise pour classer la phrase; nous supposons ici une classification binaire et décrivons donc deux états de sortie possibles.

Ce qui précède illustre l'action du filtre de 2 mots sur la matrice de phrases. Premièrement, le filtre de deux mots, représenté par la matrice jaune 2 x 5 w, recouvre les vecteurs de mots "I" et "like". Ensuite, il effectue un produit élémentaire pour tous ses 2 x 5 éléments, puis les résume et obtient un nombre (0,6 x 0,2 + 0,5 x 0,1 + ... + 0,1 x 0,1 = 0,51). 0.51 est enregistré comme premier élément de la séquence de sortie, o, pour ce filtre. Ensuite, le filtre descend d'un mot et se superpose aux mots vecteurs de "like" et "this" et effectue la même opération pour obtenir 0,53. Par conséquent, o aura la forme de (s-h + 1 x 1), dans ce cas (7-2 + 1 x 1)

Notez que la dimensionnalité de c dépend à la fois de s et de h, en d'autres termes, elle variera selon les phrases de différentes longueurs et les filtres de différentes tailles de région. Pour résoudre ce problème, les auteurs utilisent la fonction de pooling 1-max et extraient le plus grand nombre de chaque vecteur c.

# concat1max

Après 1-max pooling, nous sommes sûrs d'avoir un vecteur de longueur fixe de 6 éléments (= nombre de filtres = nombre de filtres par taille de région (2) x nombre de taille de région considéré (3)). Ce vecteur de longueur fixe peut ensuite être introduit dans une couche softmax (entièrement connectée) pour effectuer la classification. L'erreur de la classification est ensuite redéployée dans les paramètres suivants dans le cadre de l'apprentissage:

Les w matrices qui ont produit o

Le terme de biais ajouté à o pour produire c

Vecteurs de mots (facultatif, utiliser les performances de validation pour décider)

Max Pooling

Fully connected

Maintenant que nous pouvons détecter ces fonctionnalités de haut niveau, on a attaché une couche entièrement connectée à la fin du réseau. Cette couche prend essentiellement un volume d'entrée (quelle que soit la sortie de la couche conv ou ReLU ou de la couche de pool qui le précède, dans notre cas c’est la concaténation des vecteurs de caractéristique) et génère un vecteur de dimension N où N est le nombre de classes parmi lesquelles le programme doit choisir. Ici c’est le nombre de ‘serialnumber’ dans le jeu d’entrainement. Chaque nombre dans ce vecteur de dimension N représente la probabilité d'une certaine ‘serialnumber’. Par exemple, si le vecteur résultant pour un programme de classification de chiffres est [0.1 0.1 0.75 0 0 0 0 0.05], cela représente une probabilité de 10% que le texte soit ‘serialnumber 1’, une probabilité de 10% que le texte soit ‘serialnumber 2’, une probabilité de 75% que le texte soit ‘serialnumber 3’ et une probabilité de 5% que le texte soit ‘serialnumber 4’ ////(note latérale: il existe d'autres façons de représenter la sortie, mais je ne fais que montrer l'approche softmax ). La façon dont cette couche entièrement connectée fonctionne est qu'elle examine la sortie de la couche précédente (ce qui, comme nous nous en souvenons, devrait représenter les cartes d'activation des fonctionnalités de haut niveau) et détermine les caractéristiques les plus corrélées à une classe particulière. Par exemple, si le programme prédit qu'une image est un chien, elle aura des valeurs élevées dans les cartes d'activation(activation maps) qui représentent des fonctions de haut niveau comme une patte ou 4 jambes, etc. De même, si le programme prévoit qu'une image est un bird, il aura des valeurs élevées dans les cartes d’activation qui représentent des fonctionnalités de haut niveau telles que des ailes ou un bec, etc. En gros, une couche FC examine les fonctions de haut niveau les plus corrélées à une classe particulière et possède des poids particuliers. calculer les produits entre les poids et le calque précédent, vous obtenez les probabilités correctes pour les différentes classes.

# Chapitre X. Conclusions et perspectives

Tout est dans le titre : vos conclusions sur le travail effectué dans ce stage, vos conclusions sur le projet, les suites éventuelles qui en découlent.

Ce que vous avez retiré de ce stage par rapport à vos attentes, ce que vous avez apporté, …

Environ 1 à 2 pages

# Annexes

Joindre une plaquette de présentation de la société en annexe

Et toute autre annexe que vous jugez utile

# Chapitre Y. Consignes sur la rédaction (Titre niveau 1)

Attention : chaque chapitre doit commencer sur une page recto !

## Partie 1 (Titre niveau 2)

### Sous partie 1 (titre niveau 3)

Ce modèle définit des styles pour les titres (niveau 1 à 4), pour le texte (Normal en police Garamond 12), pour les légendes de figures ou de tableaux, les tables des matières, des illustrations ou des tableaux. Vous devez impérativement respecter ces styles même si vous ne les trouvez pas à votre gout.

Si vous préférez utilisez latex pour écrire votre rapport, il n’y a pas de difficulté. Utilisez dans le cas le modèle envoyé..

Si le rapport est confidentiel, pensez à le mentionner sur la page de garde. Il est inutile de le rajouter sur chaque page du document.

Le nombre de pages attendu pour le rapport de stage est compris entre 40 et 50 pages (annexes non comprise, sauf cas particuliers d’étudiants en double diplôme).

**Une attention particulière doit être portée aux respects des règles de grammaire et d’orthographe. Les éditeurs disposent de correcteurs, utilisez-les et faites-vous relire si besoin.**

### Sous partie 2 (titre niveau 3)

Ceci est une phrase. Ceci est une phrase. Ceci est une phrase. Ceci est une phrase. Ceci est une phrase. Ceci est une phrase. Ceci est une phrase. Ceci est une phrase.

Ceci est une phrase. Ceci est une phrase. Ceci est une phrase. Ceci est une phrase. Ceci est une phrase. Ceci est une phrase. Ceci est une phrase. Ceci est une phrase.

Maintenant on a des choses à dire.

#### Sous sous partie. (titre niveau 4).

Ceci n’est rien

#### Sous sous partie. (titre niveau 4).

Toujours rien

## Partie 2 (Titre niveau 2)

**Ne pas oublier de référencer les figures dans le texte et de donner une légende à chaque figure.**

Exemple : Sur la Figure 1, deux moutons dans un pré nous regardent, un petit et un grand.



Figure 1. Ceci est un mouton

Pour générer automatiquement la liste des figures, aller dans le menu Références, puis Insérer une légende

**Ne pas oublier de référencer les tableaux dans le texte et de donner une légende à chaque tableau.**

Faire de même pour les tableaux comme ci-après.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Tableau 1. Les jolis nombres

Pour faire des items :

* Voilà le premier item
* Et maintenant le second
* Et ainsi de suite

Et des listes numérotées

1. Premier item
2. Et le second
3. Et encore un autre

# Bibliographie

Doit commencer sur une page recto

# Liste des Illustrations

Doit commencer sur une page recto

[Figure 1. Ceci est un mouton 1](#_Toc397721280)