

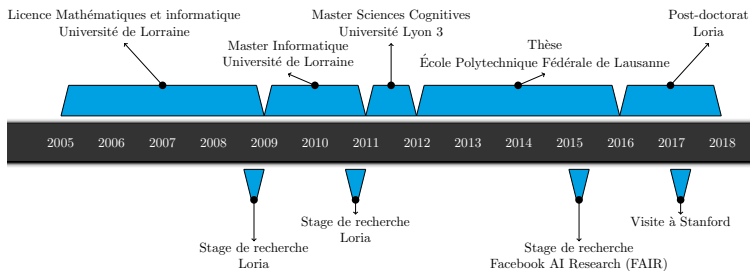
Candidature au poste de maître de conférences à
Polytech Nancy n°1087

Joël Legrand

Section 27

6 mai 2018

Formation



Thèse de doctorat :

- Soutenue en juillet 2016
- Titre : Word Sequence Modeling using Deep Learning: an End-to-end Approach and its Applications

Enseignements

Enseignements

Niveau	Formation	Intitulé	Volume horaire			
			CM	TD	TP	Total
L3	TN	Bases de Données	0	16	22	30
M1	SCA	Mise à niveau informatique	7,5	7,5	0	18
	SCA	Algorithmique pour l'Apprentissage	12	0	0	18
M2	TN	Fouilles de Données et Extraction de Connaissance	4	8	0	14
Total d'heures sur 1,5 année			23.5	31.5	22	78

TN = Télécom Nancy, SCA = Sciences de la Cognition et Applications

Encadrements

- Équipe de 3ème année (niveau M2) pour le Hackathon TELECOM nancy
- Stage de 2ème année (niveau M1) à TELECOM Nancy
- Projet tutoré de l'école des mines de Nancy

Tâches collectives

Enseignement

- Organisation de deux modules dans le cadre du Master 1 *Sciences de la Cognition et Applications*
- Participation à l'organisation du Hackathon Télécom Nancy en 2017

Recherche

- Participation au comité de lecture pour la conférence NIPS 2014 (Neural Information Processing Systems)
- Membre du comité de programme du workshop NIPS LLD 2017 (Learning with Limited Labeled Data : Weak Supervision and Beyond)

Plan

- ① CV
- ② **Projet d'enseignement**
- ③ Projet de Recherche

Projet d'enseignement

Proposition d'un cours de développement logiciel basé sur les technologies **Java EE** (Enterprise Edition) à destination des élèves de **4ème année en spécialité I2S** :

- Java EE : étend la plateforme Java SE
 - Service web, architectures distribuées, mapping objet-relationnel
- Fournit des standards permettant de :
 - Travailler à plusieurs contributeurs sur un même projet
 - Maintenir et faire évoluer une application que l'on n'a pas créée soi-même

⇒ Approche par projet en petits groupes

Projet d'enseignement

Proposition d'un cours sur l'**apprentissage automatique** à destination des élèves de **5ème année en spécialité I2S** :

- Préparation des données d'apprentissage
- Différents paradigmes d'apprentissage automatique
 - Apprentissage supervisé / non-supervisé
 - Régression / classification / clustering
- Différents modèles d'apprentissage
 - Arbres de décision, modèles markoviens, réseaux de neurones, apprentissage profond, ...
- Champ d'application et limites de l'apprentissage automatique
- Cours en anglais

Projet d'enseignement

- Autre projet d'enseignement
 - En fonction des besoins, reprise d'un cours d'**algorithmique pour l'intelligence artificielle** à destination des élèves de **4ème année en spécialité I2S**
- Responsabilités administratives
 - Au sein de la formation I2S
 - Création du contenu pédagogique pour la 5ème année Polytech
- Vie d'école
 - Hackathon
 - Coupe de France de robotique

Plan

- ① CV
- ② Projet d'enseignement
- ③ **Projet de Recherche**

Contexte de recherche

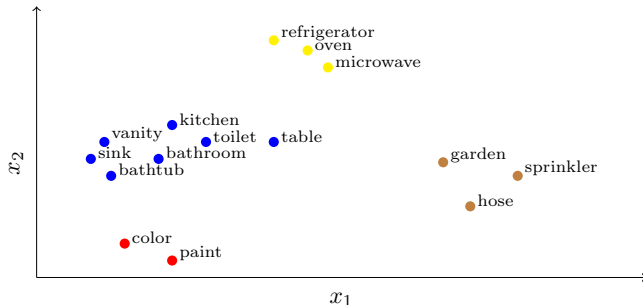
L'apprentissage profond (*deep learning*)

- Réseaux de neurones
- Généralement assez profond, mais pas forcément
- Apprentissage automatique des **représentations** à partir des données brutes (pixels, signal sonore, ...)

Contexte de recherche

L'apprentissage profond pour le traitement automatique de la langue (TAL)

- Apprentissage de représentations (*word embeddings*)
 - représentations vectorielles continues
 - de petite dimension (par rapport à la taille du dictionnaire)
 - apprises sur de larges corpus non annotés



Travail de thèse : représentations de segments de phrases

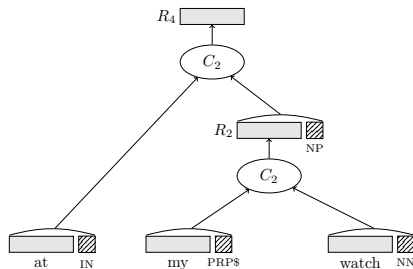
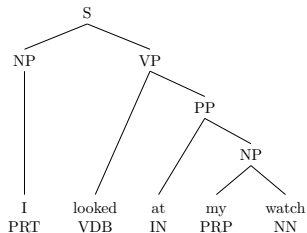
Problématiques

- Représentations vectorielles de mots
 - Obtenues à partir du contexte dans lequel les mots apparaissent
- Représentations vectorielles de segments de phrases
 - Grande diversité lexicale
 - Nombre exponentiel de formulations possibles
 - dispersion des données
 - ⇒ nécessité de combiner les représentations de mots

Objectifs de la thèse

- Étudier et proposer des représentations vectorielles de segments de phrases, dans le but de résoudre des tâches de TAL

Travail de thèse : composition de mots à l'aide de RNN



Chaque module C_i est un réseau de neurones qui

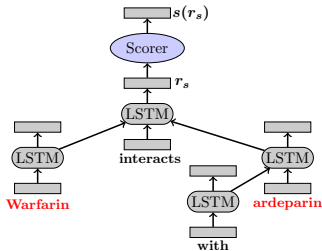
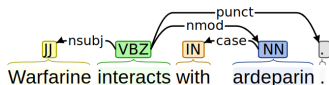
- prend i représentations continues comme entrée
- renvoie une représentation continue en sortie

Travail de thèse : Résultats

- Evalué sur PennTreebank
 - Résultats équivalents à l'analyseur syntaxique de Stanford
 - 3 fois plus rapide
 - ECML : Legrand and Collobert, 2014
 - ICLR : Legrand and Collobert, 2015
- Evalué sur les corpus SPMRL (9 langues)
 - Langues morphologiquement riches (allemand, tchèque, basque, ...)
 - Meilleure performance moyenne publiée (en terme de f1-score)
 - Meilleures performances publiées pour 6 langues
 - ACL : Legrand and Collobert, 2016

Travail de post-doctorat

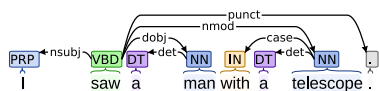
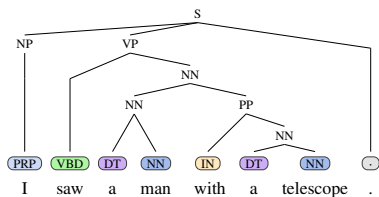
Extraction de relations entre des entités biomédicales :



- Représentation vectorielle de phrase
- RNN le long d'un arbre de dépendances

Projet de recherche : Premier axe

Analyser l'impact du **paradigme syntaxique** ainsi que celui du **modèle** choisi pour composer l'information lexicale, sur l'apprentissage de **représentations vectorielles de phrases à l'aide de RNN**.



- Évaluation quantitative sur des tâches standardisées
- Analyse qualitative (information syntaxique vs sémantique)

Projet de recherche : Deuxième axe

Étudier le potentiel des RNN pour l'apprentissage de représentations vectorielles de phrases en termes d'**apprentissage par transfert** et d'**apprentissage non et semi-supervisé**.

Apprentissage par transfert

- Transfert multilingue

Apprentissage non et semi-supervisé

- Méthodes à base d'auto-encodeurs

Conclusion

Points forts du projet d'enseignement

- Profil polyvalent pour l'enseignement de l'informatique
- Spécialisation en intelligence artificielle et apprentissage profond

Points forts du projet de recherche

- Projet de recherche fortement inscrit dans les thématiques de l'équipe Synalp
- Apport de mes compétences en apprentissage profond

Merci pour votre attention !



J. Legrand, Y. Toussaint, C. Raïssi, A. Coulet
Tree-LSTM and Cross-Corpus Training for
Extracting Biomedical Relationships from Text.
[DLPM Workshop \(ECML-PKDD\) 2017](#)



J. Legrand, M. Auli, R. Collobert
Neural Network-based Word Alignment through
Score Aggregation.
[In Proceedings of WMT \(ACL\) 2016.](#)



J. Legrand, R. Collobert
Deep Neural Networks for Syntactic Parsing of
Morphologically Rich Languages.
[In Proceedings of ACL 2016.](#)



J. Legrand, R. Collobert
Phrase Representations for Multiword
Expressions.
[In Proceedings of MWE \(ACL\) 2016.](#)



J. Legrand, R. Collobert
Joint RNN-Based Greedy Parsing and Word
Composition.
[In Proceedings of ICLR, 2015](#)



J. Legrand, R. Collobert
Reccurent Greedy Parsing with Neural
Networks.
[In Proceedings of ECML, 2014.](#)



R. Lebret, J. Legrand, R. Collobert
Is Deep Learning Really Necessary for Word
Embeddings ?.
[In NIPS Workshop on Deep Learning, 2013.](#)