



HAL
open science

Apprentissage par renforcement pour la recherche d'experts sur le web

Pegah Alizadeh, Jorge Garcia Flores, Ivan Vladimir Meza Ruiz

► **To cite this version:**

Pegah Alizadeh, Jorge Garcia Flores, Ivan Vladimir Meza Ruiz. Apprentissage par renforcement pour la recherche d'experts sur le web. EGC, Jan 2020, Brussels, Belgique. hal-02469927

HAL Id: hal-02469927

<https://hal.science/hal-02469927v1>

Submitted on 6 Feb 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Apprentissage par renforcement pour la recherche d'experts sur le web

Pegah Alizadeh*, Jorge Garcia Flores**, Ivan Vladimir Meza Ruiz***

*Léonard de Vinci Pôle Universitaire, Centre Recherche, La Défense, France
pegah.alizadeh@devinci.fr,

**LIPN-UMR CNRS 7030, PRES Sorbonne Paris Cité, Villetaneuse, France
jgflores@lipn.fr

***Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas, Mexico City
ivanvladimir@turing.iimas.unam.mx

Résumé. La recherche d'experts parmi les migrations hautement qualifiées est un enjeu crucial pour les pays en développement. Cet article présente une méthode d'apprentissage par renforcement profond pour répondre à cette problématique à partir de résultats des moteurs de recherche sur le web. Les résultats obtenus pourront être utilisés par des sociologues de la migration pour mieux comprendre les diasporas des savoirs, ainsi que par le pays en développement pour localiser ses experts formés à l'étranger. Notre méthode est basé sur des requêtes envers des moteurs de recherche pour y extraire les informations concernant l'institution et l'année d'affiliation de chaque expert. L'objectif de ce travail est de définir un navigateur intelligent capable d'assister cette recherche en générant et en observant le moins possible de requêtes automatiques. Nous utilisons comme navigateur un *Deep-Q Network* avec deux architectures basées sur des réseaux de neurones pour approximer la valeur de la Q -value fonction.

1 Introduction

Diego Martínez¹ a grandi en Argentine et a étudié dans un programme de Master en Biotechnologie à l'université de Buenos Aires. En 2010, il a obtenu une bourse internationale et il est partie à l'université d'Essex pour y faire une thèse. Où est-il maintenant ? Est-ce qu'il est resté en Angleterre ou est-il rentré en Argentine ? Où est-ce qu'il a travaillé de 2015 à 2018 ? En d'autres termes, comment a été son évolution professionnelle au cours de ces dernières années ? Localiser les membres expatriés de la diaspora hautement qualifiée est un défi majeur pour les pays en développement (Meyer et Wattiaux, 2006). Les sociologues de la migration analysent les mouvements professionnels de ces diasporas à partir de sources statistiques comme les recensements de population, les enquêtes sur la population active ainsi que des données administratives. Cependant, ces sources ne sont pas entièrement satisfaisantes et ils représente une quantité non négligeable de travail manuel (Turner et al., 2015). Nous supposons que la fouille du web pourrait être une approche plus efficace pour mieux comprendre la dynamique de ces mouvements migratoires (Flores et al., 2012).

1. Tous les noms ont été changés à fin de protéger les identités

Apprentissage par renforcement profond pour la recherche d'experts sur le web

Cet article présente un travail en cours sur la recherche de ces experts expatriés à partir du web avec une méthode basée sur l'apprentissage par renforcement profond (ARP). Notre méthode vise à récolter des informations des moteurs de recherche web afin d'extraire une liste des paires d'entités nommées (EN)s $\langle institutions, années \rangle$ liées à ces experts.

L'approche intuitive pour extraire les informations professionnelles des experts est d'interroger les moteurs de recherche web avec le nom de l'expert, enrichi par différents mots clés et puis d'y extraire les ENs sur tous les *snippets*² obtenus. L'inconvénient de cette approche est qu'elle est coûteuse à la fois en termes de mémoire et de temps de calcul. Nous proposons une méthode automatique qui sélectionne les requêtes et les *snippets* les plus prometteurs en termes de richesse de contenu. Plusieurs problèmes doivent être pris en compte si nous voulons mettre en œuvre une recherche automatique (voir Figure 1) : comment nous devrions-nous sauter d'un *snippet* à l'autre?; et comment faut-il modifier les ENs extraites après avoir traité un nouveau *snippet*? Quelles ENs sont correctement associées à un certain expert et lesquelles ne le sont pas? Pour atteindre cet objectif, nous souhaitons suivre une stratégie optimale qui génère un nombre minimum de requêtes et traite un nombre minimum de *snippets*.

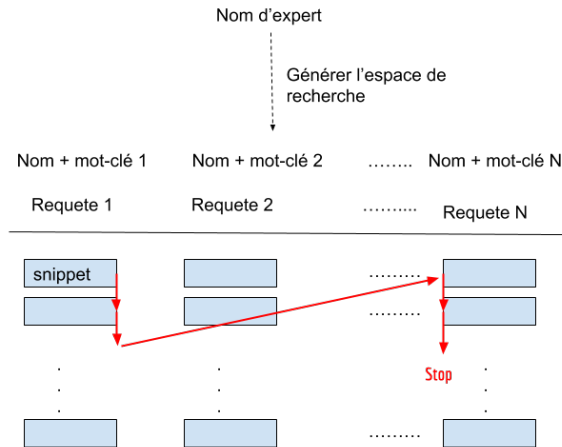


FIG. 1 – Un exemple d'espace de recherche généré pour un expert.

Des travaux récents montrent des résultats prometteurs avec des méthodes basées sur un *Deep-Q Network* pour l'extraction d'informations (Narasimhan et al., 2016), et la résolution de problèmes arithmétiques (Wang et al., 2018). Inspiré par le travail de (Narasimhan et al., 2016) sur l'extraction d'informations à partir des résultats de recherche web, nous modélisons la tâche d'extraction des informations professionnelles de l'expert comme une recherche de politique optimale dans un cadre d'apprentissage par renforcement (AR). Par rapport à leur approche, notre défi est d'apprendre un modèle avec différents espaces de recherche (à chaque expert correspond un espace de recherche unique), tandis que dans le travail de référence il y a un seul espace de recherche pour chaque base de données (*shooting* et *FoodShield*). Nous proposons deux modes de navigation "intelligente" basés sur un *Deep-Q Network* (Mnih et al., 2015). La fonction valeur de Q dépend de la séquence des *snippets* sélectionnée ainsi que de son contenu textuel. Nous utilisons aussi deux réseaux de neurones : l'un à propagation avant (NN)³ et l'autre un *Long Short Term Memory Network* (LSTM). Nous avons fait des tests sur

2. Un *snippet* est une entrée fournie par un moteur de recherche en réponse à une requête, composé d'un titre, une description brève et une adresse web.

3. feedforward neural network

une base de données contenant une liste d’experts proposée par des sociologues de la migration. Pour chaque expert, un *gold standard* (étalon-or) a été créé manuellement par des sociologues (voir Section 3). À notre connaissance, c’est la première approche d’extraction de recherche d’experts basée sur des méthodes apprentissage par renforcement profond (ARP).

Explorations et recherche des experts sur le web. La plupart des travaux récentes en recherche d’experts utilisent comme source soit l’ensemble des publications scientifiques propre à chaque expert (Buitelaar et al., 2014), soit le "profilage" sémantique de l’expert (Sateli et al., 2017). L’axe principal de ces travaux est centré sur la fouille de *topics* (sujet d’expertise), de sorte que ni l’aspect sociologique de la mobilité ni les problèmes techniques de recherche web ne sont pas pris en compte. La plupart de travaux étudient les relations expert et son sujet d’expertise (Liu et al., 2017). Le web de données est également utilisé comme une source de recherche d’experts (Stankovic et al., 2011). D’autre part, le WePS (web People Search campaigns) (Amigó et al., 2010) consistait à classifier par clustering ainsi qu’à identifier les noms de personne à partir des résultats de recherche web. Toutefois, aucun enrichissement par des mot-clés pour ces requêtes a été utilisée dans cette tâche (Artiles et al., 2009). Après le WePS, les recherches de personnes sur le web ont disparu et ses méthodes ont été appliquées au traitement de la réputation en ligne (Manaman et al., 2016) et à la désambiguïsation des EN en utilisant des profils d’utilisateur sur les réseaux sociaux (Spina et al., 2013).

Apprentissage par Renforcement Profond (ARP). Dans l’ARP, un agent interagit avec l’environnement en sélectionnant des actions dans des états. Ce processus est guidé par la fonction représentant la politique $\pi : S \rightarrow A$. La politique est calculable en attribuant une pénalisation ou une récompense $r(s, a)$ à chaque état et action et en maximisant l’espérance de la somme des récompenses (Sutton et Barto, 1998). La qualité des politiques est définie par la fonction de valeur Q ; $Q : S \times A \rightarrow \mathbb{R}$ donnée par $Q^\pi(s, a) = \mathbb{E}[\sum_i \gamma^{i-1} R_i | \pi]$. Si Q^* représente la valeur Q de la politique optimale, la *politique optimale* peut être représentée par : $\pi^*(s) = \operatorname{argmax}_a Q^{\pi^*}(s, a) \forall s \in S$. L’algorithme de *Q-learning* est une méthode indépendante du modèle pour apprendre la fonction optimale de la valeur de Q . Calculer $Q(s, a)$ pour des environnements réels avec une grande taille (ou un espace continu) des états ou actions n’est pas pratique. Pour résoudre ce problème Mnih et al. (2015) introduisent l’algorithme de *Deep-Q Newtok* (DQN) pour approximer la fonction de valeur de Q avec un réseau de neurones (RN) multicouches non linéaires. Étant donné l’état s et l’action a , le DQN donne la valeur $Q(s, a; \theta)$ où θ représente les paramètres du RN.

2 ARP pour extractions d’évolutions de carrières

Notre but est d’extraire une liste des doublons des ENS liée à chaque expert (affiliation, années) représentant sa trajectoire professionnelle pendant la période active. Pour définir un système de récompense, il est nécessaire d’avoir un *gold standard*. Dans notre cas la seule information disponible pour chaque expert est le nom de la personne, son affiliation institutionnelle, l’année de début et l’année de fin ou une sous-ensemble de telles informations. Nous proposons un modèle de Processus de Décision Markovien (*PDM*) pour répondre aux questions suivantes (voir Figure 1) : (a) Comment devrions-nous nous “déplacer” parmi les requêtes, c’est-à-dire : quand avons-nous besoin de lancer une nouvelle requête ? (b) Après avoir sélectionné une requête contenant une liste de *snippets* indexés par les moteurs de recherche, comment sélectionner ces *snippets*, y compris les bonnes ENS pour un nom d’expert donné ?

(c) Lorsque nous avons extrait des nouvelles ENs à partir d'un nouveau *snippets*, comment devrions-nous gérer l'ensemble de ENs extraites jusqu'à cette étape ? Une fois que l'EN a été extraite, faut-il l'ajouter dans la trajectoire de mobilité professionnelle ou pas ? (d) Dans quelle mesure devrions-nous continuer à analyser des *snippets* avant d'utiliser une nouvelle requête ? (e) Comment faut-il définir la fonction de récompense par rapport aux *gold standards* ?

Dans ce travail, nous générons 7 requêtes pour chaque expert de la base de données de #AIExtranjero (voir Section 3). Ces requêtes sont générées à partir des résultats de recherche du nom de la personne sur le web, et puis d'enrichir ce nom avec les mots-clés suivants *doctorate, institute, Master, PhD, undergraduate, university*. Les requêtes sont envoyées à 5 moteurs de recherche (DuckduckGo, Google, Bing, CiteSeerx et Researchgate), et les *snippets* récoltés sont stockés comme des espaces de recherche.

Processus de Décision Markovien. Nous modélisons notre navigateur dans les espaces de recherche sous la forme d'un agent de Deep-Q Network. En pratique, nous avons besoin d'une modèle de PDM pour chaque expert et son *gold standard*. Le PDM comprend les paramètres suivants :

États. Nous utilisons deux modèles d'états et deux réseaux de neurones pour calculer la fonction de la valeur Q .

1 *États à base vectorielle pour un réseau de neurones multi-couches (NN)*. Dans ce cas, chaque *snippet* est représenté par un vecteur de valeurs réelles contenant :

- Un codage *one-hot* de taille 7 indiquant quel type de requête est utilisée pour générer le *snippet*.
- Un codage *one-hot* de taille 5 indiquant le moteur de recherche.
- Un vecteur de dimension 6 contenant des statistiques sur des données communes (affiliation, années) entre le *snippet* analysé et le *gold standard*.

Exemple : Nous supposons $\mathcal{E} = \{E, Ad, Af\}$ est la trajectoire de mobilité dans le *gold standard* pour l'expert où E est son affiliation institutionnelle de référence et Ad et Af sont l'année de début et l'année de fin de référence dans le *gold standard*. $\mathcal{E}' = \{e_1, e_2, \dots, e_n, a_1, a_2, \dots, a_m\}$ est la trajectoire représentée par une liste d'institutions et des années extraite à partir d'un certain *snippet*. Les e_i s sont des institutions et les a_j s sont des années. Ensuite, le vecteur de statistique est comme suit : $\langle |\{e_i\}_i \cap \{E\}|, |\{a_j\}_j \cap \{Ad, Af\}|, |\mathcal{E} \cap \mathcal{E}'|, |\mathcal{E}|, |\mathcal{E}'|, |\mathcal{E} \cup \mathcal{E}'| \rangle^4$.

- Une valeur binaire indiquant si le nom de la personne extrait du *snippet* est une variante valide⁵ du nom de la personne à partir de la requête.

2 *États textuels encodées dans le Long Short Term Memory (LSTM)* : Chaque *snippet* est considéré comme un état d'entrée dans un modèle *Sequence to sequence*.

Actions. Mise à part l'action `stop`, nous avons deux types d'actions possibles : q et d . Les actions de type q sont : supprimer la requête actuelle (q_{DELETE}), ajouter une nouvelle requête (q_{ADD}), continuer avec la requête actuelle (q_{KEEP}). q_{ADD} choisit une requête à partir des requêtes enrichies en mots-clés mentionnées dans la section 2. Les actions de type d continuent à analyser l'ensemble de *snippets* de la requête en cours. Étant donné que l'algorithme accumule une liste d'ENs sur la trajectoire professionnelle de chaque expert, ces actions déterminent comment des nouvelles ENs extraites doivent être prises en compte. Les actions de type d peuvent :

4. le $|\cdot|$ signifie la longueur d'ensemble.

5. Par exemple : *D. Martínez* est une variante valide de *Diego Martínez*

accepter l’institution extraite du *snippet*, rejeter l’institution extraite, accepter l’année extraite du *snippet*, rejeter l’année extraite, et accepter l’institution et l’année extraits du *snippet*. Et finalement l’action de `STOP` définie quand il faut arrêter le traitement pour l’expert en cours.

Récompenses. Nous Supposons que $\mathcal{E}_i = \{e_i, a1_i, a2_i\}$ sont des ENs extraites jusqu’à la sélection de l’action d . Nous supposons également que $\mathcal{E}_{i+1} = \{e_{i+1}, a1_{i+1}, a2_{i+1}\}$ sont des ENs extraites après avoir sélectionné l’action d associée au *gold standard* $\mathcal{E} = \{E, Ad, Af\}$. Nous définissons une fonction de précision pour chaque \mathcal{E}_i par rapport aux ENs de référence dans le *gold standard* : $\text{ACC}(\mathcal{E}_i) = 1.0_{(\text{if } e_i=E)} + \sum_{j=1}^2 0.5_{(\text{if } a_{j_i} \in \{Ad, Af\})}$. La fonction de récompense $r(s, a)$ est définie comme suit : $r(s, q_{\text{ADD}}) = -10.0$; $r(s, q_{\text{DELETE}}) = 0.0$; $r(s, q_{\text{KEEP}}) = 0.0$ et $r(s, d) = -0.1 + \text{ACC}(\mathcal{E}_{i+1}) - \text{ACC}(\mathcal{E}_i)$. La fonction de précision garantit que les ENs extraites sont effectivement liées à l’expert. En ajoutant une petite valeur négative -0.1 , nous conservons le nombre des *snippets* analysés aussi bas que possible⁶.

Algorithme. Nous avons mis en oeuvre un algorithme DQN (Mnih et al., 2015) avec deux réseaux de neurones différents : un réseau multi-couches à propagation avant (NN) et un réseau LSTM. Après l’entraînement du DQN sur un sous-ensemble de notre jeu d’experts, une fonction Q optimale nous permet d’extraire l’ensemble des institutions d’affiliation ainsi que les années que l’expert a passé dans cette institution.

3 Cadre expérimental

Base de données #AIExtranjero.

La base de données contient des informations provenant de 8048 étudiants mexicains ayant reçu une bourse du gouvernement CONACYT pour étudier un master ou une thèse à l’étranger entre 1981 et 2005. Les données ont été recueillies par un sociologue de la migration et elles incluent le pays de destination, l’institution d’affiliation, le sujet de la thèse, le domaine scientifique et le début et la fin année de financement pour chaque personne. Le tableau 1 montre un exemple d’un expert avec les champs que nous avons utilisé dans nos expériences. Nous avons divisé la base de données #AIExtranjero (en résumé #AE.) en deux sous-ensembles d’entraînement et de test (voir tableau 2). Pour chaque expert de la base de données, une ensemble de sept requêtes (voir Section 2) avec cinq moteurs de recherche a été généré afin de générer un espace de recherche suffisamment grand pour chacun.

Algorithmes de référence. Nous comparons nos méthodes avec deux algorithmes :

– Méthode de référence *exhaustive* : Pour chaque expert de la base #AE., cette *baseline* traite la totalité des 7 requêtes et elle analyse tous les *snippets* de l’espace de recherche de manière

Name	Institution	Start year	End year
Diego Martínez	Essex University	2012	2015

TAB. 1 – Un extrait de trajectoire d’expert de la base de données #AIExtranjero

#AE.	Entraînement	Test
Nombre de personnes	7557	491

TAB. 2 – Nombre de personnes dans la base de données #AIExtranjero

6. La raison pour définir la valeur -10.0 est de pénaliser fortement le traitement en cas d’utilisation d’une nouvelle requête. Bien que le valeur 1.0 encourage la navigation pour trouver une nouvelle EN correctement

Apprentissage par renforcement profond pour la recherche d'experts sur le web

		Baselines			
NE	Accuracy	DQN NN	DQN LSTM	Exhaustive	Frequency
Institution	exact	0.895	0.897	0.984	0.513
	P	0.536	0.533		
	R	0.447	0.448		
	F1	0.444	0.442		
Year	exact	0.285	0.284	0.655	0.142
	P	0.123	0.118		
	R	0.168	0.168		
	F1	0.125	0.122		

TAB. 3 – Précision des DQN+NN et DQN+LSTM sur le test du jeux de données #AE.

	#AE.		#AE.	
	DQN NN	DQN LSTM	Exhaustive baseline	Frequency baseline
pourcentage des snippets analysés	24.1%	24.4%	100%	100 %
moyenne de nombre de requêtes utilisées	3.184	1.890	7	7

TAB. 4 – Proportion des snippets analysés par rapport à la moyenne des requêtes générées

exhaustive. Elle calcule la meilleure précision possible sur l'extraction des instituts et des années pour l'expert.

– Méthode de référence par *fréquence* : Cette *baseline* extrait d'abord tous les ENs de tous les *snippets*, puis elle sélectionne les ENs ayant la fréquence la plus élevée.

La méthode de référence *exhaustive* est la plus précise mais elle ne tient pas en compte les limites de mémoire et le temps d'exécution. On peut voir la deuxième méthode comme une forme de génération automatique des *gold standards*.

Paramétrage. Nous entraînons nos architectures de réseau de neurones DQN+NN (NN avec 3 couches) et DQN+LSTM (avec le réseau LSTM), pour apprendre une stratégie optimale afin d'extraire les évolutions professionnelles des experts. Nous définissons les paramètres comme le suivant : la taille de la mémoire de relecture (*replay memory*) \mathcal{D} est définie sur 2000 et le *discount factor* $\gamma = 0.95$. Pour chaque compilation de réseau, la fonction objectif (*loss function*) est définie avec l'erreur quadratique moyenne (*mse*) et un optimiseur d'Adam comme un algorithme du gradient stochastique pour faire converger les réseaux de neurones. Pour ajuster l'exploration-exploitation, nous définissons le ϵ de la stratégie ϵ -*greedy* à 0.1. Dans DQN+NN, le NN est un réseau de neurones de 3 couches cachées et une fonction d'activation RELU. Dans DQN+LSTM, le LSTM est modélisé avec un LSTM bi-directionnel de taille 64 et une couche dense de dimension 16 et une fonction d'activation linéaire.

Résultats. Les résultats montrent les performances de nos algorithmes en extraction de trajectoires sur les données de test. Nous montrons la moyenne calculée sur 5 exécutions différentes d'agent (navigateur) entraîné. Le tableau 3 compare la précision des algorithmes DQN+NN et DQN+LSTM avec les approches de base. La méthode exhaustive montre la meilleure précision possible pour l'extraction des instituts. Nous supposons $\mathcal{E}' = \{e_1, \dots, e_K, a_1, \dots, a_L\}$ sont les ENs extraites par notre algorithme pour un expert donné et $\mathcal{E} = \{E, A_1, A_2\}$ sont les valeurs de référence contenues dans le *gold standard*. La précision *exact* est : $\text{acc}(\mathcal{E}, \mathcal{E}') = 1.0_{\text{if } E \in \{e_i\}_i} + \sum_{i=1}^2 0.5_{\text{if } A_i \in \{a_k\}_k}$. Ces résultats montrent que les deux algorithmes ont de bons scores de précision pour l'extraction des institutions (0.897 v.s. 0.984). Par contre la précision est améliorable en ce qui concerne l'extraction des années (0.285 v.s. 0.655).

Le tableau 4 montre le pourcentage de *snippets* analysés par rapport à tous les *snippets* disponibles sur l'espace de recherche de chaque expert, ainsi que le nombre moyen de requêtes nécessaires. Les résultats montrent que l'algorithme DQN+LSTM est capable d'extraire des

paires (institut, années) pour les personnes du jeu de données #AE avec une précision de 0.897 en analysant que 24.4% des *snippets*, avec une moyenne de 1.890 requêtes par personne, ce qui est très proche de la précision obtenue par la base de référence exhaustive après avoir analysé 100% des *snippets* avec une moyenne de 7 requêtes par expert. La figure 2 montre l'évolution de l'extraction d'information pendant la phase d'entraînement. Comme on peut le constater, le DQN+LSTM converge plus rapidement que le DQN+NN pour les espaces de recherche de grande taille alors qu'ils ont presque la même précision à la fin.

4 Conclusions et perspectives

Dans cet article, nous avons présenté une méthode de recherche d'experts sur le web basé sur l'apprentissage profond renforcé et un algorithme de *Deep Q-Network* avec deux différentes architectures de réseaux de neurones pour l'approximation de la valeur de Q . Les résultats obtenus montrent le potentiel de notre méthode : le navigateur basé sur le DQN+LSTM est capable d'extraire des informations professionnelles avec une précision de 0.897 dans l'extraction de l'affiliation institutionnelle de l'expert, et une précision de 0.284 pour l'extraction des années de début et de fin d'affiliation, tout en analysant seulement 24.4% de la totalité des *snippets* issue des requêtes web.

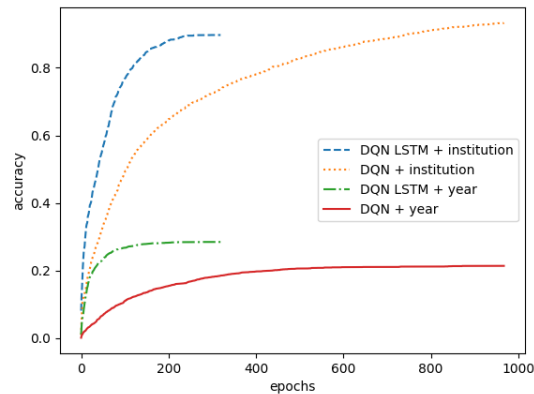


FIG. 2 – Évolution de la précision sur l'espace de recherche #AE pendant l'entraînement.

Ce projet est financé par la subvention ECOS-Nord (M15MH01) et par l'opération RENFO du Labex EFL (axe 5). Nous tenons à remercier Josue Fabricio Urbina González et Carl Theodoro Posthuma-Solis pour leur aides au développement du projet.

Références

- Amigó, E., J. Artiles, J. Gonzalo, D. Spina, B. Liu, et A. Corujo (2010). Weps3 evaluation campaign : Overview of the on-line reputation management task. In *CLEF*.
- Artiles, J., J. Gonzalo, et E. Amigó (2009). The impact of query refinement in the web people search task. In *ACL*.
- Buitelaar, P., G. Bordea, et B. Coughlan (2014). Hot topics and schisms in NLP : community and trend analysis with saffron on ACL and LREC proceedings. In *LREC*.

- Flores, J. J. G., P. Zweigenbaum, Z. Yue, et W. Turner (2012). Tracking researcher mobility on the web using snippet semantic analysis. In *ANLP*.
- Liu, J., B. Jia, H. Xu, B. Liu, D. Gao, et B. Li (2017). A topic rank based document priors model for expert finding. In *Advanced Computational Methods in Life System Modeling and Simulation*.
- Manaman, H. S., S. Jamali, et A. AleAhmad (2016). Online reputation measurement of companies based on user-generated content in online social networks. *Computers in Human Behavior*.
- Meyer, J.-B. et J.-P. Wattiaux (2006). Diaspora knowledge networks : Vanishing doubts and increasing evidence. *IJMS* 8(1).
- Mnih, V., K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, et M. A. Riedmiller (2015). Playing atari with deep reinforcement learning. *CoRR abs/1312.5602*.
- Narasimhan, K., A. Yala, et R. Barzilay (2016). Improving information extraction by acquiring external evidence with reinforcement learning. In *EMNLP*.
- Sateli, B., F. Löffler, B. König-Ries, et R. Witte (2017). Scholarlens : extracting competences from research publications for the automatic generation of semantic user profiles. *PeerJ Computer Science* 3.
- Spina, D., J. Gonzalo, et E. Amigó (2013). Discovering filter keywords for company name disambiguation in twitter. *Expert Systems with Applications*.
- Stankovic, M., J. Jovanovic, et P. Laublet (2011). Linked data metrics for flexible expert search on the open web. In G. Antoniou, M. Grobelnik, E. Simperl, B. Parsia, D. Plexousakis, P. De Leenheer, et J. Pan (Eds.), *The Semantic Web : Research and Applications*.
- Sutton, R. S. et A. G. Barto (1998). *Introduction to Reinforcement Learning* (1st ed.). Cambridge, MA, USA : MIT Press.
- Turner, W., J. Garcia Flores, et M. de Saint Leger (2015). Computer supporting diaspora knowledge networks : a case study in managing distributed collective practices. In J.-B. Meyer (Ed.), *Diaspora : towards the new frontier*. Institut de recherche pour le développement.
- Wang, L., D. Zhang, L. Gao, J. Song, L. Guo, et H. T. Shen (2018). Mathdqn : Solving arithmetic word problems via deep reinforcement learning. In *AAAI*.

Summary

The search for experts among highly qualified migrations is a crucial issue for developing countries. This article presents a deep reinforcement learning method to address this issue from search engine results on the web. The obtained results can be used by migration sociologists to better understand the diasporas of knowledge, as well as for developing countries to locate their trained experts abroad. Our method is based on queries to search engines to retrieve information about the institution and year of affiliation of each expert. The goal of our work is to define a smart browser capable of assisting this search by generating and observing as few automated queries as possible. This is done by means of a *Deep-Q network* and various neural network architectures for the Q-value function approximation.