
Régularisation Spatiale de Représentations Distribuées de Mots

Paul Mousset^{1,2} — Yoann Pitarch¹ — Lynda Tamine¹

1. IRIT, Université de Toulouse, CNRS, Toulouse, France

2. Atos Intégration, Toulouse, France

RÉSUMÉ. Stimulée par l'usage intensif des téléphones mobiles, l'exploitation conjointe des données textuelles et des données spatiales présentes dans les objets spatio-textuels (p. ex. tweets) est devenue la pierre angulaire à de nombreuses applications comme la recherche de lieux d'attraction. Du point de vue scientifique, ces tâches reposent de façon critique sur la représentation d'objets spatiaux et la définition de fonctions d'appariement entre ces objets. Dans cet article, nous nous intéressons au problème de représentation de ces objets. Plus spécifiquement, confortés par le succès des représentations distribuées basées sur les approches neuronales, nous proposons de régulariser les représentations distribuées de mots (c.-à-d. plongements lexicaux ou word embeddings), pouvant être combinées pour construire des représentations d'objets, grâce à leurs répartitions spatiales. L'objectif sous-jacent est de révéler d'éventuelles relations sémantiques locales entre mots ainsi que la multiplicité des sens d'un même mot. Les expérimentations basées sur une tâche de recherche d'information qui consiste à retourner le lieu physique faisant l'objet (sujet) d'un géo-texte montrent que l'intégration notre méthode de régularisation spatiale de représentations distribuées de mots dans un modèle d'appariement de base permet d'obtenir des améliorations significatives par rapport aux modèles de référence.

ABSTRACT. Stimulated by the heavy use of smartphones, the joint use of textual and spatial data in space-textual objects (eg., tweets) has become the mainstay of many applications, such as the finding of places of interest. These tasks are fundamentally based on the representation of spatial objects and the definition of matching functions. In this article, we focus on the representation of these objects. More precisely, reinforced by the success of distributed word representations approaches, we propose to regularize word embeddings that can be combined to construct object representations, using their spatial distributions. The purpose is to reveal possible local semantic relationships between words and the multiplicity of meanings of the same word. Experiments based on a semantic location prediction task demonstrate that the integration of our method of spatial retrofitting of word embeddings into a basic matching model provides significant improvements over strong baselines.

MOTS-CLÉS : plongement lexical, apprentissage hors-ligne, géo-texte

KEYWORDS: word embeddings, retrofitting, geo-text

1. Introduction

Au cours des dernières décennies, la création de contenus géoréférencés, aussi connus également sous le nom de données spatio-textuelles ou géo-textes, a fortement augmenté. Parmi les données spatio-textuelles, nous pouvons notamment citer : les lieux d'intérêts (c.-à-d. *POI*¹) et leur description textuelle, les publications géolocalisées (p. ex. les tweets), les photos géotaggées et leurs mots-clefs (p. ex. les photos Instagram) ainsi que les pointages (c.-à-d. *check-ins*) via des services géodépendants (p. ex. Foursquare). L'interaction entre le texte et la localisation géographique offre des perspectives intéressantes pour un large éventail d'applications telles que la gestion de crise (Imran *et al.*, 2015) et les guides touristiques (Deveaud *et al.*, 2015). Cette prévalence soulève également d'importantes questions de recherche sous-jacentes à l'appariement d'objets spatio-textuels, qui est l'étape clef de diverses tâches telles que l'interrogation de géo-textes (Zhang *et al.*, 2014), la mention de lieux (Fang et Chang, 2014 ; Han *et al.*, 2018) ou la prédiction sémantique de lieux (Dalvi *et al.*, 2009 ; Zhao *et al.*, 2016). Les solutions existantes pour l'appariement d'objets spatio-textuels reposent principalement sur l'utilisation d'une combinaison d'éléments textuels et spatiaux pour construire des représentations efficaces d'objets (Zhang *et al.*, 2014) ou pour définir des modèles efficaces d'appariement objet-objet (Dalvi *et al.*, 2009 ; Zhao *et al.*, 2016). Notre travail s'inscrit précisément dans la perspective d'apprentissage de représentations idoines de géo-textes partant des représentations des granules de ces objets, à savoir les mots, et de leurs répartitions spatiales.

Plus précisément, compte tenu du succès des représentations distribuées des mots basées sur les approches neuronales, appelées également plongements lexicaux, l'objectif du travail présenté dans cet article est d'étendre ces représentations avec les connaissances issues des répartitions spatiales des mots représentés. Le but de cette extension est de construire des représentations améliorées qui pourraient être utilisées efficacement tant pour la représentation d'objets que pour leur appariement. Nous basons nos réflexions sur deux hypothèses principales : (1) les cooccurrences de paires de mots dans des objets spatio-textuels révèlent des relations sémantiques locales entre ces mots. Par exemple, les mots *dinosaure* et *musée* sont sémantiquement très reliés près d'un musée d'histoire naturelle, mais moins reliés près d'un musée d'art ; (2) comme corollaire à l'hypothèse 1, des sens distincts d'un même mot pourraient être identifiés en utilisant la répartition spatiale des mots comme source d'évidence. Par exemple, le mot *dinosaure* peut se référer en général à un animal préhistorique ou à une chaîne de restaurants dans le contexte local de New York. Ainsi, nous exploitons la répartition spatiale des mots pour identifier des relations sémantiques locales entre mots ainsi que des sens locaux de mots. Nous résumons dans ce qui suit les contributions présentées dans cet article :

– Une stratégie de régularisation a posteriori (c.-à-d. *retrofitting*) (Faruqui *et al.*, 2014 ; Vulić et Mrkšić, 2018) comme moyen d'affiner des représentations distribuées

¹. Point Of Interest

de mots (c.-à-d. *word-embeddings*) à l'aide de répartitions spatiales.

– Une validation empirique de nos hypothèses de recherche et une étude expérimentale montrant l'efficacité des plongements lexicaux spatiaux que nous proposons d'intégrer dans une tâche de prédiction sémantique du lieu.

L'article est organisé comme suit : la Section 2 donne un aperçu des travaux connexes. La Section 3 présente les notions de base ainsi que les hypothèses de recherche. La Section 4 décrit en détail la méthodologie proposée pour la régularisation a posteriori des représentations distribuées des mots à l'aide des connaissances spatiales. Les résultats de l'évaluation expérimentale sont présentés et discutés en Section 5. Enfin, la Section 6 conclut l'article et présente quelques perspectives.

2. Méthodes pour la régularisation de plongements lexicaux

L'apprentissage de représentations distribuées de mots, appelées également plongements lexicaux (c.-à-d. *word embeddings*), a été très étudiée ces dernières années, notamment depuis l'émergence d'algorithmes populaires tels que *Word2vec* (Mikolov *et al.*, 2013b), *GloVe* (Pennington *et al.*, 2014) ou *FastText* (Bojanowski *et al.*, 2017). Cependant, les plongements lexicaux présentent quelques inconvénients dont principalement : (1) ils ne permettent pas de lever le problème de polysémie puisque les différents sens d'un même mot sont regroupés en un seul vecteur ; (2) les vecteurs de représentations distribuées de mots peuvent s'avérer peu lisibles puisqu'ils ne sont pas alignables avec des ressources externes (Nguyen *et al.*, 2018).

Pour répondre au problème de polysémie, l'approche généralement employée est d'apprendre conjointement les mots et leurs sens en utilisant des ressources sémantiques externes (Cheng *et al.*, 2015 ; Iacobacci *et al.*, 2017 ; Mancini *et al.*, 2017). Par exemple, Iacobacci *et al.* (2017) commence par désambiguïser les mots à l'aide de la ressource *Baberfy*, puis révisé la fonction objectif du *CBOW* (Mikolov *et al.*, 2013b) pour apprendre conjointement les représentations des mots et des sens. Cheng *et al.* (2015) étendent le modèle *Word2vec* en identifiant des paires concepts-mots dans un contexte donné et apprennent le mot et sa signification dans le même espace latent.

Une approche standard pour améliorer le problème de manque de lisibilité des représentations distribuées consiste à injecter des connaissances externes, principalement des contraintes de ressources lexicales, en utilisant soit une approche *en ligne*, soit une approche *hors-ligne* (Mikolov *et al.*, 2013b ; Pennington *et al.*, 2014). L'approche en ligne vise à régulariser la fonction objectif en exploitant des connaissances externes pendant l'étape d'apprentissage (Xu *et al.*, 2014 ; Yu et Dredze, 2014 ; Nguyen *et al.*, 2017). Par exemple, Yu et Dredze (2014) et Xu *et al.* (2014) proposent le Modèle à Contrainte Relationnelle (*Relation Constrained Model*) qui étend la fonction objectif du modèle *Skip-Gram* (Mikolov *et al.*, 2013b) avec une relation sémantique entre les paires de mots, telle que fournie par une ressource lexicale, basée sur l'hypothèse que les mots apparentés produisent des contextes similaires. L'approche hors-ligne, également appelée régularisation a posteriori (*retrofitting*), utilise des res-

sources externes en dehors de l'étape d'apprentissage pour affiner les plongements lexicaux (Faruqui *et al.*, 2014 ; Mrkšić *et al.*, 2016 ; Vulić et Mrkšić, 2018 ; Nguyen *et al.*, 2018). Par exemple, Faruqui *et al.* (2014) proposent une méthode pour améliorer les représentations distribuées des mots en privilégiant les mots apparentés, tels que fournis par une ressource lexicale (p. ex. *WordNet*, *FramNet*), pour avoir une représentation vectorielle similaire. Mrkšić *et al.* (2016) et Vulić et Mrkšić (2018) ont quant à eux basé leurs travaux sur une approche d'attraction et de répulsion. L'hypothèse sous-jacente est de modéliser une fonction objectif d'apprentissage visant à rapprocher les mots similaires (synonymes) et à rejeter les mots dissemblables (antonymes).

En l'état courant de nos connaissances, notre travail est la première tentative de correction spatiale de représentations distribuées de mots et de discrimination des sens locaux des mots à base de répartitions spatiales.

3. Préliminaires

Nous introduisons dans cette section les notions de base abordées dans cet article permettant de répondre à notre problématique de recherche. Nous commençons par définir les objets spatio-textuels et les propriétés qui en découlent. Nous énonçons ensuite les hypothèses de recherche puis définissons formellement le problème abordé.

3.1. Définitions

Pour rappel, l'objectif de notre travail est de construire des représentations de mots augmentées par des connaissances issues des répartitions spatiales. Cela induit l'utilisation de ressources géotaggées, appelées objets spatio-textuels, mais aussi de concepts spatiaux telles que la distance et la répartition spatiale, que nous définissons ci-après.

Définition 1. (Objet spatio-textuel) Un objet spatio-textuel o est un texte géotaggé (p. ex. un lieu d'intérêt avec une description textuelle). Le géotag est représenté par une paire de coordonnées (*latitude*, *longitude*) se référant à l'emplacement géographique l , noté $o.l$ (p. ex. la localisation physique d'un lieu). Chaque objet o est décrit par un ensemble d'attributs textuels, valorisés ou non selon la nature de l'objet, incluant par exemple son nom $o.nom$, une courte description $o.des$ ou un historique des critiques/remarques émises par les utilisateurs *p.cr.* Ainsi, un lieu d'intérêt (p. ex. issu de Google Place² ou du réseau social Foursquare³), contient l'ensemble des attributs décrits précédemment, tandis qu'un tweet géotaggé (issu du réseau social Twitter⁴) ne contient que les attributs $o.l$, contenant les coordonnées de son lieu de publication, et $o.des$, contenant le texte rédigé par l'utilisateur. Nous adoptons une représentation

2. <https://developers.google.com/places/web-service/>

3. <https://foursquare.com>

4. <https://twitter.com>

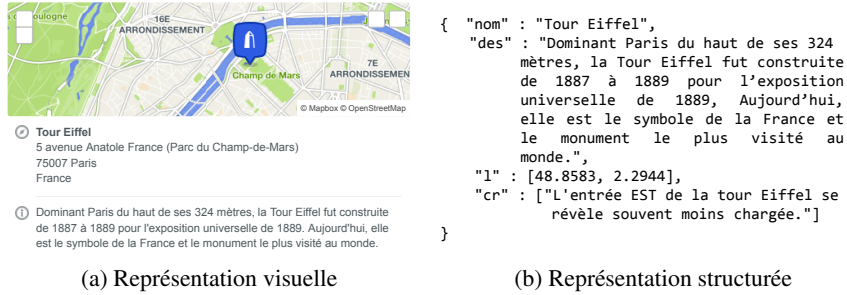


Figure 1 : Exemple d'un objet spatio-textuel issu de Foursquare.

vectorielle de l'objet o basée sur ses attributs textuels $o = [w_1^{(o)}, \dots, w_m^{(o)}]$, où chaque mot $w_i^{(o)}$ est issu d'un vocabulaire \mathcal{W} .

Exemple 1. (Objet spatio-textuel) La Figure 1 présente un objet spatio-textuel décrivant la *Tour Eiffel* de Paris, tel qu'il apparaît sur Foursquare (Figure 1a) et sous sa forme structurée (Figure 1b). À partir des attributs textuels de l'objet (nom, description et commentaires des utilisateurs), nous obtenons sa représentation vectorielle $o = ['tour', 'eiffel', 'dominant', 'paris', \dots, 'entrée', 'est', 'tour', \dots, 'chargée']$.

Définition 2. (Distance spatiale entre des objets spatio-textuels) La distance spatiale d entre les objets spatio-textuels o_i et o_j fait référence à la distance géographique, basée sur une mesure de distance (p. ex. Haversine⁵, Euclidienne, etc.), notée $dist$, entre les emplacements $o_i.l$ et $o_j.l$. Formellement, $d(o_i, o_j) = dist(o_i.l, o_j.l)$.

L'objet d'étude de ce travail étant l'unité lexicale, c.-à-d. le mot, nous définissons à présent les concepts de répartition spatiale d'un mot et de distance spatiale entre mots.

Définition 3. (Répartition spatiale d'un mot) L'ensemble $O_i = \{o_{i,1}, o_{i,2}, \dots\}$ associé au mot w_i correspond à l'ensemble des objets spatio-textuels $o_{i,k}$ citant le mot w_i . La répartition spatiale d'un mot w_i , notée S_i , est donc représentée par l'ensemble des coordonnées $o_{i,k}.l$ des objets $o_{i,k}$ contenus dans l'ensemble O_i .

Définition 4. (Distance spatiale entre des mots) La distance spatiale d entre les mots w_i et w_j qualifie la proximité spatiale entre leur répartition spatiale respective. Plus formellement, soient les mots w_i et w_j et S_i et S_j leur répartition spatiale respective. La distance entre les mots w_i et w_j , notée $d(w_i, w_j)$, est telle que :

$$d(w_i, w_j) = dist(\phi(S_i), \phi(S_j))$$

où $\phi(S_i)$ est une fonction qui agrège la répartition spatiale de w_i en une localisation. Par exemple, $\phi(S_i)$ peut retourner les coordonnées du barycentre de S_i .

5. https://fr.wikipedia.org/wiki/Formule_de_haversine

3.2. Hypothèses

La régularisation des plongements lexicaux par des répartitions spatiales repose sur les deux hypothèses suivantes :

Hypothèse 1. Les mots dont les répartitions spatiales sont similaires ont tendance à être reliés sémantiquement. Plus précisément, plus les mots sont spatialement proches au sens de la distance spatiale (cf. Définition 4), c.-à-d. avec une distance $d(w_i, w_j)$ relativement faible, plus leurs vecteurs de représentations distribuées doivent être similaires au sens de la distance géométrique (p. ex. cosinus). Intuitivement, le mot *coupe* est sémantiquement plus relié au mot *football* en Europe qu’aux États-Unis.

Hypothèse 2. Considérons le sens local d’un mot comme étant représenté par l’ensemble des mots spatialement proches, comme énoncé par l’hypothèse 1. Un mot peut véhiculer des sens locaux différents selon la zone géographique où il est spatialement dense. En pratique, cela se traduirait par des zones particulièrement denses dans la répartition spatiale d’un mot. Intuitivement, *football* en Europe ne référence pas le même sport aux États-Unis au vu de la différence des mots spatialement proches à chacun.

3.3. Définition du problème

En nous appuyant sur l’hypothèse 1, nous supposons que les signaux issus des répartitions spatiales des mots pourraient contribuer à la construction de représentations distribuées de vecteurs de mots. Pour cela, nous proposons d’injecter ces signaux dans le processus d’apprentissage de ces représentations. Cette direction est motivée par de précédents travaux (Faruqui *et al.*, 2014 ; Vulić et Mrkšić, 2018) qui utilisent des ressources lexicales (p. ex. WordNet, FrameNet) pour contraindre l’apprentissage des représentations des mots. Cependant, les représentations distribuées, même corrigées ne permettent pas de distinguer les différents sens d’un même mot puisqu’ils sont regroupés en un seul vecteur (Mancini *et al.*, 2017 ; Iacobacci *et al.*, 2017). En d’autres termes, à chaque mot, ne correspond qu’un seul vecteur. Aussi, en tenant compte de l’hypothèse 2, nous souhaitons construire, pour chaque mot, un ensemble de représentations distribuées basées sur ses statistiques d’occurrence au sein des objets spatio-textuels associés. Formellement, à partir d’un ensemble de représentations vectorielles de mots $\widehat{\mathbf{W}} = \{\widehat{\mathbf{w}}_1, \dots, \widehat{\mathbf{w}}_n\}$, où $\widehat{\mathbf{w}}_i$ est le vecteur k -dimensionnel construit pour un mot cible $w_i \in \mathcal{W}$ en utilisant un modèle linguistique neuronal standard (p. ex. le modèle *Skip-gram* (Mikolov *et al.*, 2013b)), le problème est de déterminer pour chaque mot w_i , l’ensemble des mots spatiaux associés $\widehat{\mathbf{w}}_i^s = \{\widehat{\mathbf{w}}_{i,1}^s, \dots, \widehat{\mathbf{w}}_{i,j}^s, \dots, \widehat{\mathbf{w}}_{i,n_i}^s\}$. Chaque vecteur de mot spatial $\widehat{\mathbf{w}}_{i,j}^s$, issu d’une représentation distribuée $\widehat{\mathbf{w}}_i$ initiale, désigne la représentation distributionnelle localisée du mot w_i sur une zone spatiale dense, et n_i est le nombre de sens localisés distincts du mot w_i provenant de sa répartition spatiale sur les objets spatio-textuels O_i auquel il appartient.

4. Méthodologie

L'algorithme de construction des plongements lexicaux spatiaux est décrit par l'Algorithme 1. La première étape (lignes 1 et 2) consiste à identifier les « empreintes spatiales » de chacun des mots, c'est à dire les zones spatiales où ils sont denses. La seconde étape (lignes 3 à 5) permet de corriger les plongements lexicaux des mots en tenant compte des répartitions spatiales. Nous détaillons dans les sections suivantes chacune des étapes du processus.

4.1. Partitionnement spatial

Pour chaque mot w_i , nous commençons par identifier l'ensemble des objets spatio-textuels O_i auxquels il appartient (ligne 1). Pour identifier les zones spatiales denses du mot w_i , nous effectuons un partitionnement en k -moyennes (MacQueen *et al.*, 1967) (ligne 2). Formellement, pour chaque mot w_i , l'algorithme des k -moyennes partitionne l'ensemble des coordonnées $O_{i,l}$ issues des objets O_i en k groupes spatiaux en minimisant la distance entre les points à l'intérieur de chacune des partitions. Les k groupes spatiaux sont représentés par leur barycentre $\mathcal{B}_i = \{\mathcal{B}_{i,1}, \dots, \mathcal{B}_{i,n_i}\}$, où $\mathcal{B}_{i,j}$ est le j -ième barycentre du mot w_i et n_i le nombre

Algorithme 1 : Construction de plongements lexicaux spatiaux

Entrées : Vocabulaire \mathcal{W} ; Ensemble de représentations distribuées de mots

$\widehat{\mathbf{W}} = \{\widehat{\mathbf{w}}_1, \dots, \widehat{\mathbf{w}}_{|\mathcal{W}|}\}$; Ensemble d'objets spatio-textuels O ;

Rayon des mots proches r^+ ; Rayon des mots distants r^-

Sorties : Ensemble de représentations distribuées de mots

$\mathbf{W}^s = \{\mathbf{w}_{1,1}^s, \dots, \mathbf{w}_{1,n_1}^s, \dots, \mathbf{w}_{|\mathcal{W}|,1}^s, \dots, \mathbf{w}_{|\mathcal{W}|,n_k}^s\}$

// Partitionnement spatial (Sect. 4.1)

pour $i \in \{1, \dots, |\mathcal{W}|\}$ **faire**

1 | $O_i = \text{ExtraireObjets}(w_i, O)$

2 | $\text{PartitionnementSpatial}(O_i, \mathcal{B}_i, n_i)$

fin

// Plongements lexicaux spatiaux (Sect. 4.2)

répéter

pour $i \in \{1, \dots, |\mathcal{W}|\}$ **faire**

pour $j \in \{1, \dots, n_i\}$ **faire**

3 | $W_{i,j}^+ = \text{Voisins}(w_i, \mathcal{B}_{i,j}, \mathcal{B}, r^+)$

4 | $W_{i,j}^- = \text{Distants}(w_i, \mathcal{W}, \mathcal{B}_{i,j}, \mathcal{B}, r^-)$

5 | $\mathbf{w}_{i,j}^s = \text{Régularisation}(\widehat{\mathbf{w}}_i, W_{i,j}^+, W_{i,j}^-)$

fin

fin

jusqu'à Convergence;

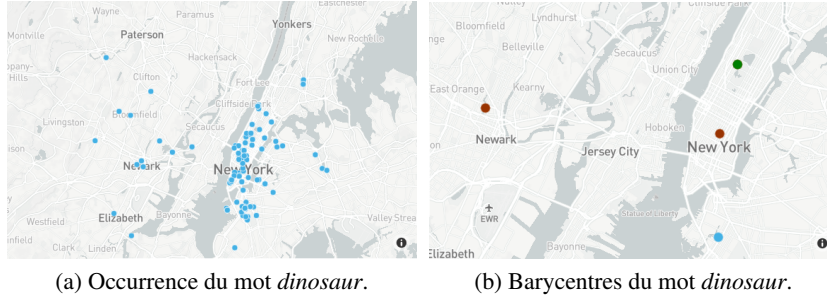


Figure 2 : Exemple du partitionnement en k -moyennes appliqué au mot *dinosaur*.

optimal de groupes pour le mot w_i déterminé en utilisant l'analyse de la silhouette (Rousseeuw, 1987). Chaque barycentre $\mathcal{B}_{i,j}$ peut être vu comme un représentant spatial de la zone qui donne lieu à un sens local du mot w_i représenté par le vecteur distribué $\mathbf{w}_{i,j}^s$.

Exemple 2. (Partitionnement spatial du mot *dinosaur*) Un exemple de partitionnement spatial est présenté en Figure 2 pour le mot *dinosaur* dans la ville de New York et ses environs. Chaque point sur la Figure 2a correspond à une occurrence du mot *dinosaur*. L'application de l'algorithme de partitionnement nous permet d'obtenir quatre barycentres, chacun révélant une spécificité locale du mot *dinosaur*. Comme le révèle la Figure 2b, chacun des points représente une spécificité locale du mot *dinosaur* : les points marrons sont plutôt associés à une chaîne de restauration populaire à New York, *Dinosaur Bar-B-Que*, qui possède notamment des établissements dans le quartier de Manhattan et la ville de Newark, tandis que le point vert est plutôt associé au *American Museum of Natural History*, un musée consacré entre autres aux dinosaures.

4.2. Représentations distribuées de mots intégrant des contraintes spatiales

Nous détaillons à présent l'étape principale de construction des représentations distribuées spatiales des mots $\mathbf{w}_{i,j}^s$, basée sur un processus de régularisation a posteriori du vecteur distribué $\widehat{\mathbf{w}}_i$ en considérant à la fois les mots voisins $W_{i,j}^+$ et les mots éloignés $W_{i,j}^-$ du barycentre $\mathcal{B}_{i,j}$.

Notre objectif est d'apprendre l'ensemble des représentations distribuées spatiales des mots \mathbf{W}^s , autrement dit, de corriger les plongements lexicaux préalablement entraînés (ligne 5). Nous contraignons que le vecteur de représentation $\mathbf{w}_{i,j}^s$ soit :

- 1) géométriquement proche (selon une métrique de distance), c.-à-d. sémantiquement relié, de la représentation distribuée du mot $\widehat{\mathbf{w}}_i$;
- 2) géométriquement proche de ses voisins spatiaux $W_{i,j}^+$;

3) géométriquement non proche des mots spatialement distants $W_{i,j}^-$.

Pour cela, nous devons déterminer les ensembles de mots voisins et de mots distants pour chacun des barycentres des mots $B_{i,j}$. L'Algorithme 2 détaille les fonctions *Voisins* et *Distants* utilisées. Pour rechercher l'ensemble des mots proches $W_{i,j}^+$ du barycentre $B_{i,j}$, nous parcourons, pour chaque mot w_k candidat (ligne 3), ses barycentres $B_{k,l}$ (ligne 4). Dans le cas où la distance séparant le barycentre $B_{i,j}$ d'un des barycentres $B_{k,l}$ est inférieure à un seuil r^+ (lignes 5), le mot w_k associé à ce barycentre est ajouté à l'ensemble $W_{i,j}^+$. De même, pour rechercher l'ensemble des mots distants $W_{i,j}^-$, nous commençons par déterminer l'ensemble des mots voisins situés dans un rayon r^- autour barycentre $B_{i,j}$ (ligne 8). L'ensemble des mots distants est donc constitué des mots qui ne sont pas contenus dans l'ensemble des mots voisins.

Algorithme 2 : Recherche des ensembles de mots voisins et distants

Entrées : Barycentre $B_{i,j}$; Vocabulaire \mathcal{W} ; Ensemble des barycentres \mathcal{B} ;
Rayon des mots proches r^+ ; Rayon des mots distants r^-

Sorties : Ensemble des mots voisins $W_{i,j}^+$; Ensemble des mots distants $W_{i,j}^-$

Fonction *Voisins*($w_i, \mathcal{W}, B_{i,j}, \mathcal{B}, r^+$)

```

1 |  $\mathcal{W}_{candidats} = \mathcal{W} \setminus w_i$ 
2 |  $voisins = \{\}$ 
3 | pour  $k \in \{1, \dots, |\mathcal{W}_{candidats}|\}$  faire
4 |   | pour  $l \in \{1, \dots, n_i\}$  faire
5 |     |   | si  $dist(B_{i,j}, B_{k,l}) < r^+$  alors
6 |       |   |   |  $voisins = voisins \cup w_k$ 
7 |     |   | fin
8 |   | fin
9 | retourner  $voisins$ 
10| fin
```

Fonction *Distants*($w_i, \mathcal{W}, B_{i,j}, \mathcal{B}, r^-$)

```

8 |  $voisins = \text{Voisins}(w_i, \mathcal{B}_{i,j}, \mathcal{B}, r^-)$ 
9 |  $distants = \mathcal{W} \setminus voisins$ 
10| retourner  $distants$ 
11| fin
```

Pour répondre à notre objectif de rapprocher sémantiquement les mots proches spatialement et d'éloigner sémantiquement les mots distants spatialement, nous proposons de minimiser la fonction objectif suivante :

$$\Psi(\mathbf{W}^s) = \sum_{i=1}^{|\mathcal{W}|} \sum_{j=1}^{n_i} \left[\alpha d(\mathbf{w}_{i,j}^s, \widehat{\mathbf{w}}_i) + \beta \sum_{w_k \in W_{i,j}^+} d(\mathbf{w}_{i,j}^s, \widehat{\mathbf{w}}_k) + \gamma \sum_{w_k \in W_{i,j}^-} 1 - d(\mathbf{w}_{i,j}^s, \widehat{\mathbf{w}}_k) \right]$$

où $d(\mathbf{w}_i, \mathbf{w}_j) = 1 - \text{sim}(\mathbf{w}_i, \mathbf{w}_j)$ est une mesure de distance dérivée de la mesure de similarité du cosinus, $\text{sim}(\mathbf{w}_i, \mathbf{w}_j) = \frac{\mathbf{w}_i \cdot \mathbf{w}_j}{\|\mathbf{w}_i\| \cdot \|\mathbf{w}_j\|}$ est la similarité du cosinus, $W_{i,j}^+$ (resp. $W_{i,j}^-$) est l'ensemble des mots spatialement **proches** (resp. **distants**) du mot $w_{i,j}$, c.-à-d. les mots situés dans un (resp. éloignés d'un) rayon r^+ (resp. r^-) autour de son barycentre $\mathcal{B}_{i,j}$, et $\alpha, \beta, \gamma \geq 0$ sont des hyperparamètres qui contrôlent l'importance relative de chaque terme. Il convient de noter que les mots compris entre les rayons r^+ et r^- autour du barycentre ne sont pas considérés pour régulariser les représentations distribuées de mots puisque considérés comme non discriminants. La fonction objectif Ψ étant différentiable, nous la minimisons par une descente de gradient.

5. Évaluation expérimentale

L'objectif de notre évaluation expérimentale est double : 1) valider les deux hypothèses formulées dans la Section 3.2 et 2) évaluer l'effet extrinsèque de la régularisation spatiale des représentations distribuées des mots au travers l'efficacité d'une tâche de recherche d'information basée sur un modèle d'appariement qui les exploite. Nous commençons par présenter la tâche d'évaluation et les jeux de données utilisés. Nous donnons ensuite quelques détails sur l'implémentation de la méthode de régularisation proposée. Enfin, nous détaillons les modèles de l'état-de-l'art permettant de résoudre la tâche ainsi que les métriques utilisées pour évaluer la qualité de nos représentations.

5.1. Tâche d'évaluation et jeux de données

Pour mener l'évaluation expérimentale, nous considérons une tâche de recherche d'information qui consiste à associer à un géo-texte, ici un tweet, le lieu dont il fait l'objet. Cette tâche est appelée prédiction sémantique du lieu (Dalvi *et al.*, 2009 ; Zhao *et al.*, 2016). Étant donné le tweet t , la tâche consiste à identifier le lieu d'intérêt p (POI) sur lequel le tweet t se concentre sémantiquement (c.-à-d. le lieu évoqué). Formellement, la tâche permet d'identifier l'unique lieu d'intérêt p , qui est l'élément $p^* \in \mathcal{P}$ le mieux classé parmi une liste ordonnée de lieux d'intérêts proposée par une fonction de correspondance sémantique $\text{Scoring}(t, p)$. La fonction Scoring est une fonction permettant de calculer un score entre le tweet t et le lieu d'intérêt p .

Pour réaliser cette tâche, nous utilisons une collection de tweets anglais géotaggés, publiée par Zhao *et al.* (2016), contenant les coordonnées géographiques des lieux où les tweets ont été émis. La collection contient plus de 590 000 tweets publiés dans les villes de New York (USA) et Singapour entre Septembre 2010 et Janvier 2015. Comme nous nous intéressons à la tâche de prédiction sémantique du lieu, nous ne retenons que les 74 000 tweets reliés à des lieux d'intérêts. En utilisant les web-services de Foursquare, nous avons collecté près de 800 000 lieux d'intérêts localisés dans les villes de New York et Singapour, ainsi que 2 millions d'avis déposés par les utilisateurs.

5.2. Détails d'implémentation

L'objet de cet article étant de régulariser des plongements lexicaux à l'aide de ré-partitions spatiales, il convient de disposer de vecteurs pré-entraînés. Pour cela, nous avons utilisé le modèle *Skip-gram* proposé par Mikolov *et al.* (2013b), qui a pour principe, étant donné un mot, prédire les mots contextuels. Les vecteurs de taille 300 ont été appris grâce à un corpus de documents issus de Wikipedia⁶ et des textes contenus dans les tweets, les lieux d'intérêts et les avis utilisateurs décrits dans la Section 5.1. Le vocabulaire contient 3.2 millions de mots qui apparaissent au moins 3 fois dans le corpus. La qualité de nos plongements lexicaux initiaux a été évaluée sur le test d'analogie Google⁷ (Mikolov *et al.*, 2013a) et a obtenu un score d'exactitude (*accuracy*) de 70% (contre 77% pour le modèle pré-entraîné de Google⁸).

Dans le cadre de l'expérimentation, nous n'effectuons le partitionnement spatial via la méthode des k -moyennes (Section 4.1) que sur les mots qui ocurrent dans les jeux de données décrits précédemment (Section 5.1), ce qui représente 238 369 mots distincts. Après le partitionnement, nous avons déterminé 630 732 groupes de mots spatiaux, soit environ 2.6 sens locaux $w_{i,j}^s$ créés par mot w_i . Nous remarquons que 166 139 (69.7%) des mots n'ont qu'une seule empreinte spatiale (c.-à-d. une seule partition spatiale $w_{i,j}$).

Nous minimisons la fonction objectif Ψ par une descente de gradient, en utilisant l'algorithme d'optimisation Adam (Kingma et Ba, 2014). Pour nos expérimentations, nous avons fixé respectivement r^+ et r^- à 100 et 500 mètres et α , β , et γ à 1.

5.3. Modèles de référence et scénarios d'évaluation

– **Modèles de référence** : pour l'évaluation comparative des performances de la tâche de prédiction sémantique de lieux, nous utilisons des modèles d'appariement issus de l'état-de-l'art. Ces modèles sont basés sur la distance, le texte, ou une combinaison des deux.

1) **DIST** (De Smith et Goodchild, 2007) : la distance Haversine entre le Tweet et lieu d'intérêt candidat;

2) **BM25** (Robertson et Jones, 1976) : le modèle probabiliste classique *BM25*;

3) **CLASS** (Zhao *et al.*, 2016) : Un modèle d'ordonnement de lieux d'intérêts qui combine la distance spatiale avec un modèle de langue :

$$Scoring(t, p) \propto \exp \left\{ -\frac{d(t, l, p, l)^2}{2\sigma^2} \right\} \times \prod_{w \in t} P(w|p)$$

avec $P(w|p)$ le modèle de langue du lieu d'intérêt, et σ^2 la variance empirique de la distance d .

6. <https://dumps.wikimedia.org/>

7. [https://aclweb.org/aclwiki/Google_analogy_test_set_\(State_of_the_art\)](https://aclweb.org/aclwiki/Google_analogy_test_set_(State_of_the_art))

8. <https://code.google.com/archive/p/word2vec/>

– **Scénarios d'évaluation** : afin d'évaluer l'effet de l'utilisation des représentations distribuées proposées, nous les injectons dans le modèle d'appariement CLASS cité ci-haut. Notre choix est motivé par le fait qu'il comporte une composante d'appariement de mots ($\prod_{w \in t} P(w|p)$) qui permet une injection aisée de représentations distribuées des objets spatio-textuels associés avec de surcroît, deux procédés possibles : ordonnancement et expansion. Plus précisément, pour un objet spatio-textuel $o = [w_1^{(o)}, \dots, w_m^{(o)}]$ donné (tweet t ou lieu d'intérêt p), nous déterminons tout d'abord sa représentation distribuée \hat{o} comme la moyenne des représentations distribuées des mots $w_i^{(o)}$ qui le composent, pondérée par l>IDF. Nous injectons ensuite la représentation distribuée \hat{o} dans la fonction d'estimation du score de pertinence selon les procédés suivants :

a) **CLASS-MATCH (CM)** (Zhao *et al.*, 2016) : c'est une configuration qui concerne la révision du calcul du score d'appariement en remplaçant le score donné par le modèle de langue ($\prod_{w \in t} P(w|p)$) par la similarité du cosinus de la paire de représentations distribuées (\hat{t}, \hat{p})

$$Scoring(t, p) \propto \exp \left\{ -\frac{d(t.l, p.l)^2}{2\sigma^2} \right\} \times sim(\hat{t}, \hat{p})$$

b) **CLASS-ETENDU (CE)** (Zamani et Croft, 2016) : c'est une configuration qui concerne l'expansion préalable du tweet t avec des mots pertinents. Nous suivons l'approche d'expansion de requête proposée par Zamani et Croft (2016). Celle-ci consiste à réécrire le tweet en exploitant la proximité des mots dans l'espace des représentations distribuées.

$$Scoring(t, p) \propto \exp \left\{ -\frac{d(t.l, p.l)^2}{2\sigma^2} \right\} \times \prod_{w \in t^*} P(w|p)$$

avec t^* le tweet étendu. Les mots étendus sont sélectionnés comme suit : (1) nous déterminons un ensemble de mots candidats $V \in \mathcal{W}$, obtenu à partir des mots contenus dans une liste ordonnée de lieux d'intérêts classés par le modèle CLASS pour le tweet t ; (2) nous calculons une distribution probabiliste en utilisant l'espace sémantique du tweet. Formellement, la probabilité de chaque terme w , étant donné la représentation distribuée \hat{t} du tweet, est calculée par $p(w|\hat{t}) = \frac{sim(w, \hat{t})}{Z}$. Le facteur de normalisation Z est déterminé par $Z = \sum_{w' \in V} sim(w', \hat{t})$; (3) nous considérons les top m termes ayant la probabilité $p(w|\hat{t})$ la plus élevée, pour étendre le tweet.

Pour chacun des deux scénarios (CM et CE) présentés ci-dessus, nous évaluons deux configurations : 1) avec les représentations distribuées classiques (CM- \widehat{W} et CE- \widehat{W}); 2) avec les représentations distribuées régularisées (CM- \mathbf{W}^s et CE- \mathbf{W}^s). Pour les scénarios qui utilisent les représentations régularisées, nous utilisons le mot local $w_i^{(t)}$ (resp. $w_i^{(p)}$) le plus proche en minimisant la distance Haversine entre l'emplacement du tweet (resp. lieu d'intérêt) et le barycentre du mot $\mathcal{B}_{i,j}$.

5.4. Métriques d'évaluation

Pour évaluer les performances des modèles de référence et scénarios, nous appliquons la technique de ré-ordonnement. Pour chaque tweet t de la collection, les 200 lieux d'intérêts les plus proches spatialement sont sélectionnés et retenus comme candidats potentiels ($Acc@200 = 100\%$). L'ordonnement via les modèles de référence se base donc sur cette liste.

Nous exploitons deux mesures d'évaluation connues : la précision (ou *accuracy*) notée $Acc@k$ (Powers, 2011) et le rang réciproque moyen (ou *Mean Reciprocal Rank*), noté MRR (Craswell, 2009). Étant donné la tâche de prédiction sémantique du lieu, il est important de mentionner que de faibles valeurs de k sont particulièrement considérées.

6. Résultats expérimentaux

Dans cette section, nous nous intéressons d'une part, à la validation des hypothèses proposées, et d'autre part, à l'évaluation des performances de nos représentations distribuées et régularisées dans le cadre la tâche de prédiction sémantique de lieu, décrite précédemment.

6.1. Validation des hypothèses

Pour valider les hypothèses détaillées dans la Section 3.2, nous construisons les cartes thermiques (*heat map*) des valeurs de similarité du cosinus entre les représentations distribuées d'un échantillon de mots. Afin de représenter les sens locaux des mots, nous divisons notre collection de données en deux sous-ensembles distincts, dé-

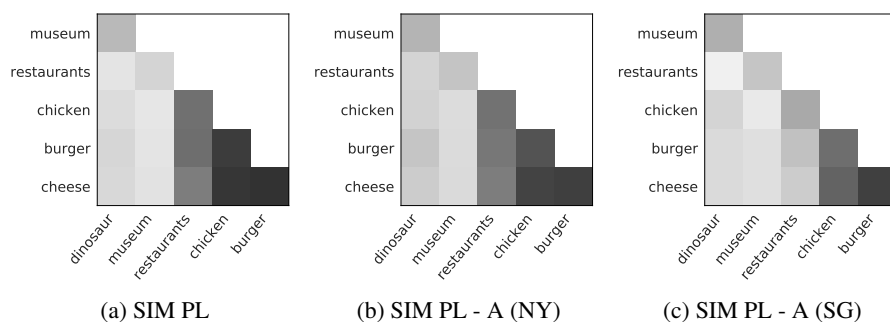


Figure 3 : Similarité du cosinus des plongements lexicaux (SIM PL) traditionnels (a), amortis (SIM PL - A) par les distances des barycentres des mots dans les sous-ensembles de New York (NY) (b) et de Singapour (SG) (c)

terminés selon la ville dont ils sont originaires (c.-à-d. à New York ou à Singapour). Pour chaque mot de chaque ville, nous déterminons son unique barycentre correspondant à sa zone d'apparition moyenne. Enfin, pour chaque paire de mots (w_i, w_j) , les similarités du cosinus sont atténuées par un facteur spatial $f_s(w_i, w_j)$ qui permet de tenir compte de leur proximité spatiale. Formellement, $f_s(w_i, w_j)$ est défini par $f_s(w_i, w_j) = \exp\{-\frac{d(\mathcal{B}_i, \mathcal{B}_j) - \mu}{\sigma}\}$ avec $d(\mathcal{B}_i, \mathcal{B}_j)$ la distance Haversine entre les barycentres de w_i et w_j et μ (resp. σ) la distance moyenne (resp. l'écart type) entre toutes les paires de mots d'un sous-ensemble.

Les résultats sont présentés dans la Figure 3. Plus la cellule est sombre, plus la paire de mots est similaire (c.-à-d. à une valeur de similarité du cosinus élevée). La Figure 3a montre la carte thermique non-pondérée (c.-à-d. similarité simple) tandis que les Figures 3b et 3c montrent respectivement les cartes thermiques pondérées par le facteur spatial pour les villes de New York et Singapour. Nous pouvons par exemple remarquer que la cellule (*restaurants, dinosaur*) est plus sombre dans la Figure 3b que dans Figure 3a tandis qu'elle est plus claire dans Figure 3c que dans Figure 3a. D'une manière générale, il n'y a pas de raison évidente pour que les mots *restaurants* et *dinosaur* soient liés l'un à l'autre, comme le montre la similarité de leurs plongements lexicaux dans la Figure 3a. Cependant, nous avons noté l'existence d'une chaîne de restaurants à New York, appelée *Dinosaur Bar-B-Que*, ce qui conduit à une forte proportion d'objets géo-textuels où les deux termes cooccurrent, conduisant ainsi à une relation sémantique locale plus forte, pour cette paire de mots, dans la ville de New York, comme le révèle la Figure 3b. Ce constat permet d'étayer notre *hypothèse 1*. Par ailleurs, l'observation conjointe de la Figure 3a et de ses variations spatiales en Figures 3b et 3c fournit quelques indices sur le bien fondé de l'*hypothèse 2*. En effet, nous pouvons constater que les mots *dinosaur* et *museum* sont similaires quel que soit l'endroit. En reliant cette observation à la précédente, nous pouvons en déduire que le mot *dinosaur* pourrait se référer tant à *museum* qu'à *restaurant* dans la ville de New York, comme le révèle la forte similarité avec des mots tels que *burger* et *cheese* dans Figure 3b qui est nettement moins prononcée dans Figure 3c. Ce raisonnement vaut également pour d'autres paires de mots telles que *burger* et *restaurants*.

En résumé, cette analyse qualitative démontre que 1) la similarité sémantique entre deux mots dépend de la distance spatiale entre les sources de ces mots (c'est-à-dire que plus les sources sont proches, plus les termes sont sémantiquement reliés) et 2) que le mot peut avoir des sens distincts selon sa répartition spatiale au travers des objets où il occure. Cela motive la nécessité d'apprendre des représentations distribuées distinctes pour un mot donné en fonction de la zone géographique des objets qui le référencent.

6.2. Évaluation de l'efficacité de la régularisation spatiale des représentations distribuées des mots

Nous analysons maintenant la qualité des représentations distribuées régularisées via la tâche de prédiction sémantique du lieu décrite dans la section 5.1. Les résultats

	Modèle	<i>MRR</i>		<i>Acc@1</i>		<i>Acc@5</i>	
		Valeur	%Chg	Valeur	%Chg	Valeur	%Chg
Modèles de référence	DIST	0.514	+17.51 *	0.430	+19.61 *	0.605	+15.45 *
	BM25	0.423	+42.79 *	0.307	+67.68 *	0.668	+4.49 *
	CLASS	0.507	+19.13 *	0.401	+28.43 *	0.624	+11.79 *
Scénarios basés sur les représentations classiques	CM- \widehat{W}	0.521	+15.93 *	0.413	+24.52 *	0.640	+9.06 *
	CE- \widehat{W}	0.563	+7.28 *	0.470	+9.51 *	0.659	+5.94 *
Scénarios basés sur les représentations régularisées	CM- W^s	0.577	+4.68 *	0.489	+5.32 *	0.675	+3.36 *
	CE- W^s	0.604	–	0.515	–	0.698	–

Tableau 1 : Évaluation de l’effet de la régularisation spatiale des représentations distribuées des mots. %Chg : Taux d’amélioration relatif entre le modèle CE- W^s et les autres modèles. Test t de Student’ * : $p < 0.05$.

sont résumés dans le Tableau 1. Nous présentons les valeurs obtenues pour chacune des métriques ainsi que les taux d’améliorations relatifs (%Chg).

Dans l’ensemble, nous remarquons que les scénarios impliquant l’appariement avec les représentations distribuées régularisées (CM- W^s et CE- W^s) dépassent largement tous les modèles de référence. Par exemple, CE- W^s affiche de meilleurs résultats en termes de *MRR* avec des accroissements relatifs compris entre 19.13% et 42.79% par rapport aux modèles DIST, BM25 et CLASS. Plus précisément, le scénario CE- W^s permet un appariement tweet - lieu d’intérêt plus efficace : plus de 48% des tweets sont associés à leur lieu d’intérêt correspondant dès les premiers résultats (c.-à-d. *Acc@1*), contre 43% pour DIST. De plus, nous pouvons observer que l’injection de représentations distribuées (classiques ou régularisées) permet d’améliorer l’efficacité du modèle CLASS. Par exemple, la mesure *MRR* du scénario CE- W^s (resp. CE- \widehat{W}) augmente significativement de 16.06% (resp. 9.95%) par rapport au modèle CLASS.

En examinant plus particulièrement les deux scénarios impliquant des représentations distribuées régularisées (c.-à-d. CM- W^s et CE- W^s), nous pouvons remarquer que le scénario CE- W^s améliore la *MRR* de 4.68% et la mesure *Acc@1* de 5.32% comparé au scénario CM- W^s . La mesure *Acc@1* (resp. *MRR*) est passée de 0.489 (resp. 0.577) à 0.515 (resp. 0.604). Ces résultats pourraient s’expliquer par l’approche utilisée pour injecter les plongements lexicaux spatiaux. Tandis que dans le scénario CE- W^s , les vecteurs sont utilisés de façon indépendante pour étendre la description du tweet avant l’étape d’appariement, ils sont plutôt agrégés dans le scénario CM- W^s , pour construire des représentations de tweets et de lieux d’intérêts pondérées par l’IDF, ce qui génère un biais de représentation. Cette observation montre clairement l’impact positif de l’utilisation intrinsèque des représentations distribuées régularisées spatialement.

Pour conclure l’analyse de la qualité des représentations distribuées régularisées des mots, nous illustrons dans la Figure 4, la carte thermique des similarités entre un

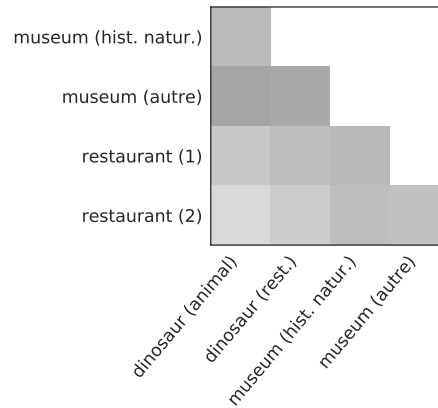


Figure 4 : Similarité du cosinus des représentations distribuées régularisées, pour les mots *dinosaur* (animal et chaîne de restauration), *museum* (musée d’histoire naturelle et un autre musée) et *restaurant* (deux types de restaurants).

sous-ensemble des mots issus de la Figure 3. Plus précisément, nous déclinons dans cette figure les sens locaux et régularisés que nous avons identifié grâce à l’Algorithme 1. Par exemple, nous avons identifié deux sens distincts du mot *museum* selon sa répartition spatiale. Nous pouvons constater, que le sens du mot *dinosaur (rest.)* est plus proche des représentations des restaurants (*restaurant (1) et (2)*) que ne l’est celui de *dinosaur (animal)*. Ceci montre bien qualitativement le rapprochement entre le résultat de la régularisation proposée et nos hypothèses initialement énoncées.

7. Conclusion

Dans cet article nous avons introduit la notion de régularisation spatiale de représentations distribuées de mots. La méthode de régularisation sous-jacente s’appliquant a posteriori (ou *retrofitting*) sur des représentations pré-entraînées s’appuie sur des répartitions spatiales pour révéler des relations sémantiques locales entre mots, ainsi que des sens distincts et locaux des mots. Nous avons évalué expérimentalement l’effet de l’utilisation des représentations régularisées proposées dans une tâche de recherche d’information, en l’occurrence la tâche de prédiction sémantique de lieux. Les résultats montrent que la méthode proposée permet d’améliorer significativement les performances de recherche comparativement à des modèles de référence de l’état de l’art. Dans le cadre de nos travaux futurs, nous envisageons de poursuivre l’évaluation de la régularisation spatiale proposée dans d’autres tâches de recherche d’information sensibles à la localisation comme le résumé spatio-temporel de tweets.

Remerciements

Cette étude a été soutenue par l'IRIT et Atos Intégration dans le cadre du programme de recherche ANRT CIFRE numéro 2016/403.

8. Bibliographie

- Bojanowski P., Grave E., Joulin A., Mikolov T., « Enriching Word Vectors with Subword Information », *TACL*, vol. 5, p. 135-146, 2017.
- Cheng J., Wang Z., Wen J.-R., Yan J., Chen Z., « Contextual text understanding in distributional semantic space », *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*, CIKM '15, p. 133-142, 2015.
- Craswell N., « Mean reciprocal rank », *Encyclopedia of Database Systems*, p. 1703-1703, 2009.
- Dalvi N., Kumar R., Pang B., Tomkins A., « A Translation Model for Matching Reviews to Objects », *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, CIKM '09, p. 167-176, 2009.
- De Smith M., Goodchild M. F., *Geospatial analysis : a comprehensive guide to principles, techniques and software tools*, 2007.
- Deveaud R., Albakour M.-D., Macdonald C., Ounis I., « Experiments with a Venue-Centric Model for Personalised and Time-Aware Venue Suggestion », *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*, CIKM '15, p. 53-62, 2015.
- Fang Y., Chang M.-W., « Entity linking on microblogs with spatial and temporal signals », *IEEE Transactions of the Association for Computational Linguistics*, vol. 2, p. 259-272, 2014.
- Faruqui M., Dodge J., Jauhar S. K., Dyer C., Hovy E., Smith N. A., « Retrofitting Word Vectors to Semantic Lexicons », *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*, NAACL-HLT '14, p. 1606-1615, 2014.
- Han J., Sun A., Cong G., Zhao W. X., Ji Z., Phan M. C., « Linking Fine-Grained Locations in User Comments », *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, n° 1, p. 59-72, 2018.
- Iacobacci I., Pilehvar M. T., Navigli R., « SenseEmbed : Learning sense embeddings for word and relational similarity », *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing*, ACL-IJCNL '17, p. 95-105, 2017.
- Imran M., Castillo C., Diaz F., Vieweg S., « Processing Social Media Messages in Mass Emergency : A Survey », *ACM Computing Surveys*, vol. 47, n° 4, p. 67 :1-67 :38, June, 2015.
- Kingma D. P., Ba J., « Adam : A method for stochastic optimization », *arXiv preprint*, 2014.
- MacQueen J. *et al.*, « Some methods for classification and analysis of multivariate observations », *Proceedings of the 5th Berkeley symposium on mathematical statistics and probability*, BSMSP 1967, p. 281-297, 1967.
- Mancini M., Camacho-Collados J., Iacobacci I., Navigli R., « Embedding Words and Senses Together via Joint Knowledge-Enhanced Training », *Proceedings of the 21st Conference on Computational Natural Language Learning*, CoNLL '17, p. 100-111, 2017.

- Mikolov T., Chen K., Corrado G., Dean J., « Efficient estimation of word representations in vector space », *arXiv preprint*, 2013a.
- Mikolov T., Sutskever I., Chen K., Corrado G. S., Dean J., « Distributed representations of words and phrases and their compositionality », *Proceedings of 27th Annual Conference on Neural Information Processing Systems 2013*, NIPS '13, p. 3111-3119, 2013b.
- Mrkšić N., Séaghdha D. O., Thomson B., Gašić M., Rojas-Barahona L., Su P.-H., Vandyke D., Wen T.-H., Young S., « Counter-fitting word vectors to linguistic constraints », *arXiv preprint*, 2016.
- Nguyen G.-H., Tamine L., Soulier L., Souf N., « Learning Concept-Driven Document Embeddings for Medical Information Search », *Proceedings of the 16th Conference on Artificial Intelligence in Medicine*, AIME '17, p. 160-170, 2017.
- Nguyen G.-H., Tamine L., Soulier L., Souf N., « A Tri-Partite Neural Document Language Model for Semantic Information Retrieval », *Proceedings of the 15th International Conference European Semantic Web Conference*, ESWC '18, p. 445-461, 2018.
- Pennington J., Socher R., Manning C. D., « Glove : Global Vectors for Word Representation. », *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, EMNLP '14, p. 1532-1543, 2014.
- Powers D. M., « Evaluation : from precision, recall and F-measure to ROC, informedness, markedness & correlation », *Journal of Machine Learning Technologies*, 37-63, 2011.
- Robertson S. E., Jones K. S., « Relevance weighting of search terms », *Journal of the American Society for Information science*, vol. 27, n° 3, p. 129-146, 1976.
- Rousseeuw P. J., « Silhouettes : a graphical aid to the interpretation and validation of cluster analysis », *Journal of computational and applied mathematics*, vol. 20, p. 53-65, 1987.
- Vulić I., Mrkšić N., « Specialising Word Vectors for Lexical Entailment », *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*, NAACL-HLT '18, p. 1134-1145, 2018.
- Xu C., Bai Y., Bian J., Gao B., Wang G., Liu X., Liu T.-Y., « Rc-net : A general framework for incorporating knowledge into word representations », *Proceedings of the 23rd ACM International on Conference on Information and Knowledge Management*, CIKM '14, p. 1219-1228, 2014.
- Yu M., Dredze M., « Improving lexical embeddings with semantic knowledge », *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics*, ACL '14, p. 545-550, 2014.
- Zamani H., Croft W. B., « Estimating Embedding Vectors for Queries », *Proceedings of the 2016 ACM on International Conference on the Theory of Information Retrieval*, ICTIR '16, p. 123-132, 2016.
- Zhang D., Chan C.-Y., Tan K.-L., « Processing Spatial Keyword Query As a Top-k Aggregation Query », *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval*, SIGIR '14, p. 355-364, 2014.
- Zhao K., Cong G., Sun A., « Annotating Points of Interest with Geo-tagged Tweets », *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, CIKM '16, p. 417-426, 2016.