



APPRENTISSAGE DE REPRÉSENTATION: DE LA DÉCISION NON-LINÉAIRE À LA GÉNÉRATION DE DONNÉES

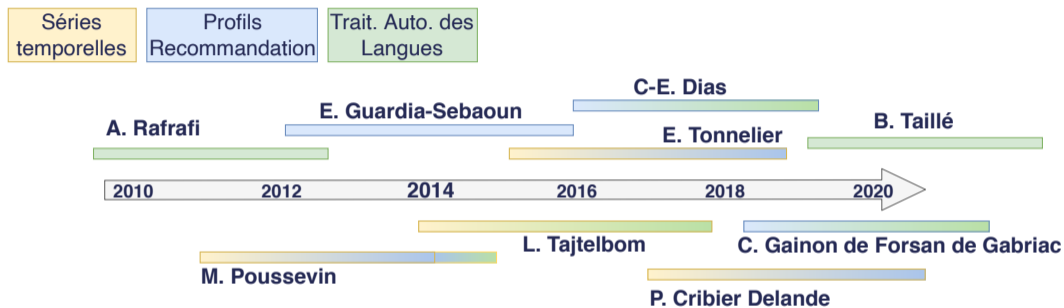
17 février 2021

Vincent Guigue

| | | | |
|--------------|-------------------|---------------------------|------------|
| Jury: | Massih-Reza Amini | Université Grenoble Alpes | Rapporteur |
| | Patrice Bellot | Université Aix-Marseille | Rapporteur |
| | Gabriella Pasi | Université de Milan | Rapporteur |
| | Mohamed Chetouani | Sorbonne Université | Examineur |
| | Martin Trepanier | Polytechnique Montréal | Examineur |
| | Emmanuel Viennet | Université Paris 13 | Examineur |



Collaborations & positionnement thématique



- Co-encadrements avec P. Gallinari, T. Artières, N. Baskiotis, L. Denoyer
- Période de mutation du machine learning & impact sociétal important
- Recherche appliquée / frontière entre modalités

(1) Sémantique *statique* des données

Sémantique = notion diffuse : plusieurs définitions
différentes échelles / diverses modalités de données

- pLSA
- Factorisation matricielle
- Word2Vec

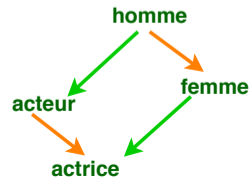
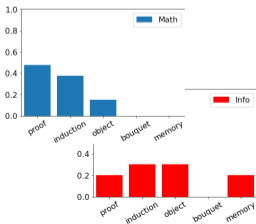
[Hofmann 99]

[Hoyer 02]

[Mikolov 13]

⇒ Representation learning

[Bengio 13]



(2) Sémantique *contextuelle* – approches génératives de bout en bout

■ Sequence 2 sequence

■ Plasticité des architectures de deep-learning

■ VAE/GAN & démêlage

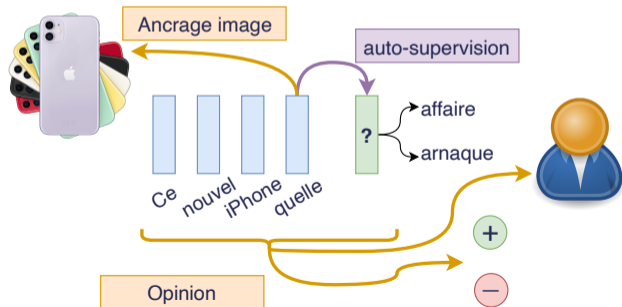
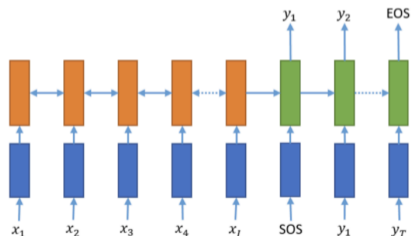
■ Modèles de langue : ELMO, BERT

[Sutskever 14]

[Antol 15]

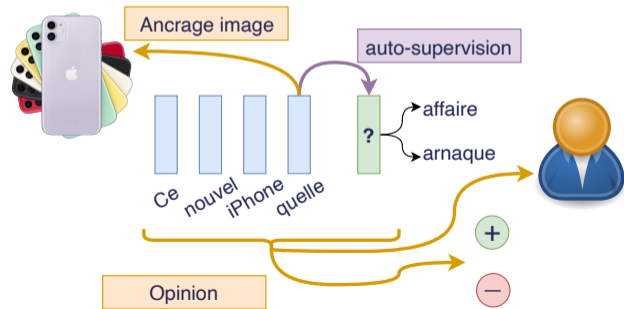
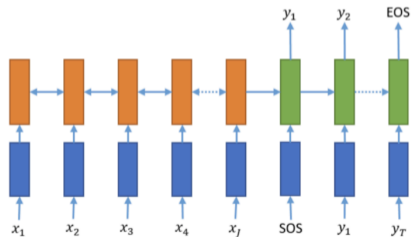
[Kingma 14, Goodfellow 14, Chen 16]

[Peters 18, Devlin 18]



(2) Sémantique *contextuelle* – approches génératives de bout en bout

- Sequence 2 sequence [Sutskever 14]
- **Plasticité des architectures** de deep-learning [Antol 15]
- VAE/GAN & démêlage [Kingma 14, Goodfellow 14, Chen 16]
- Modèles de langue : ELMO, BERT [Peters 18, Devlin 18]



(2) Sémantique *contextuelle* – approches génératives de bout en bout

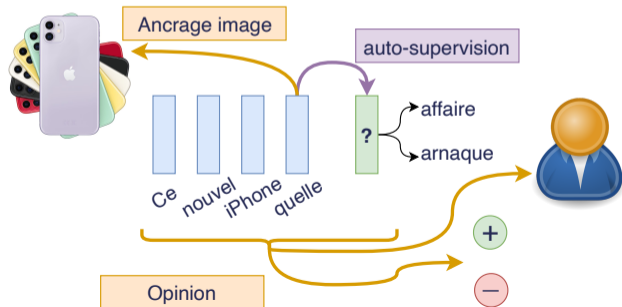
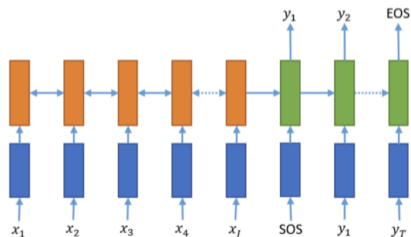
- Sequence 2 sequence
- Plasticité des architectures de deep-learning
- **VAE/GAN & démêlage**
- Modèles de langue : ELMO, BERT

[Sutskever 14]

[Antol 15]

[Kingma 14, Goodfellow 14, Chen 16]

[Peters 18, Devlin 18]



Compréhension des données,
sémantique des concepts

Sémantique : définition(s) et échelles

Différentes définitions à différentes échelles :

- Catégoriser des données : Sens = ensemble de données homogène [Dubes 88]
- LSA/pLSA (textes) Cluster = Extraction d'un champ lexical [Hofmann 99]
- Recommandation Métrique utilisateurs–produits & Affinités [Hoyer 02]
- Séries temporelles Séparation des sources, décomposition [Cordoso 97]
- Word2Vec (textes) Distance+**direction** entre les mots [Mikolov 13]
- Prod2vec Métrique fine entre les produits [Grbovic 15]

Sémantique : définition(s) et échelles

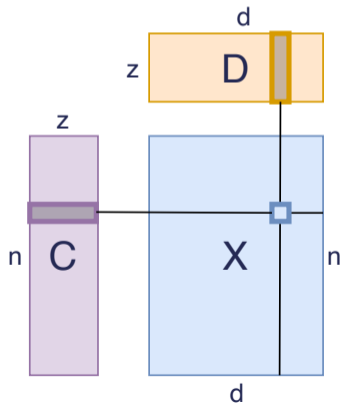
- LSA/pLSA (textes)
- Recommandation
- Séries temporelles

Cluster = Extraction d'un champ lexical [Hofmann 99]
Métrique utilisateurs–produits & Affinités [Hoyer 02]
Séparation des sources, décomposition [Cordoso 97]

Factorisation matricielle

Algorithme versatile, robuste,
débruite les données, compresse et explique

- Apprentissage par descente de gradient itérative
- Problème non convexe
- Opportunité de biaiser / contraindre l'apprentissage
 - non-negative, orthogonal



Sémantique : définition(s) et échelles

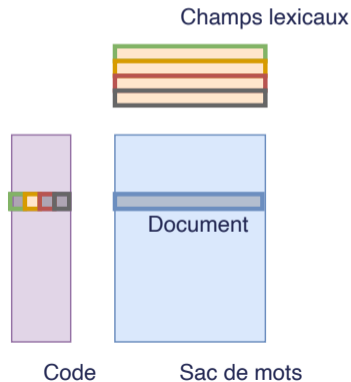
- LSA/pLSA (textes)
- Recommandation
- Séries temporelles

Cluster = Extraction d'un champ lexical [Hofmann 99]
Métrique utilisateurs–produits & Affinités [Hoyer 02]
Séparation des sources, décomposition [Cordoso 97]

Factorisation matricielle

Algorithme versatile, robuste,
débruite les données, compresse et explique

- Apprentissage par descente de gradient itérative
- Problème non convexe
- Opportunité de biaiser / contraindre l'apprentissage
 - non-negative, orthogonal



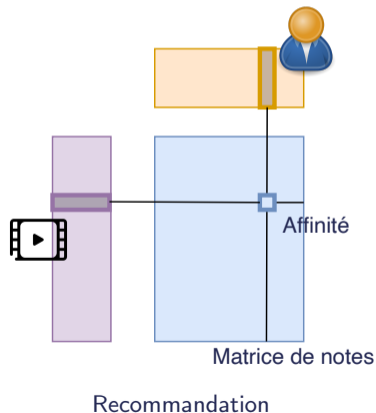
Sémantique : définition(s) et échelles

- LSA/pLSA (textes) Cluster = Extraction d'un champ lexical [Hofmann 99]
- Recommandation Métrique utilisateurs–produits & Affinités [Hoyer 02]
- Séries temporelles Séparation des sources, décomposition [Cordoso 97]

Factorisation matricielle

Algorithme versatile, robuste,
débruite les données, compresse et explique

- Apprentissage par descente de gradient itérative
- Problème non convexe
- Opportunité de biaiser / contraindre l'apprentissage
 - non-negative, orthogonal



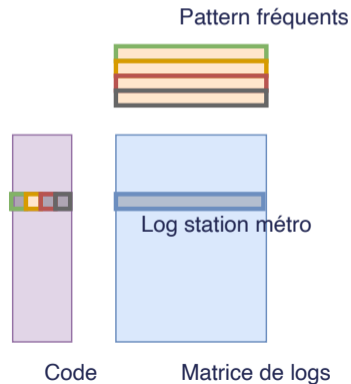
Sémantique : définition(s) et échelles

- LSA/pLSA (textes) Cluster = Extraction d'un champ lexical [Hofmann 99]
- Recommandation Métrique utilisateurs–produits & Affinités [Hoyer 02]
- Séries temporelles Séparation des sources, décomposition [Cordoso 97]

Factorisation matricielle

Algorithme versatile, robuste,
débruite les données, compresse et explique

- Apprentissage par descente de gradient itérative
- Problème non convexe
- Opportunité de biaiser / contraindre l'apprentissage
 - non-negative, orthogonal



Sémantique : définition(s) et échelles

Différentes définitions à différentes échelles :

- Word2Vec (textes)

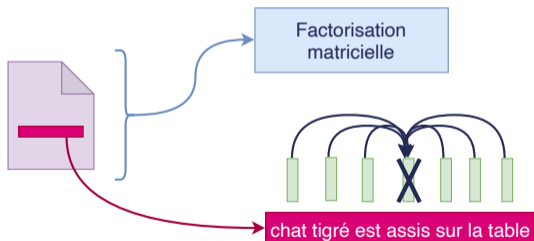
- Prod2vec

Distance+**direction** entre les mots

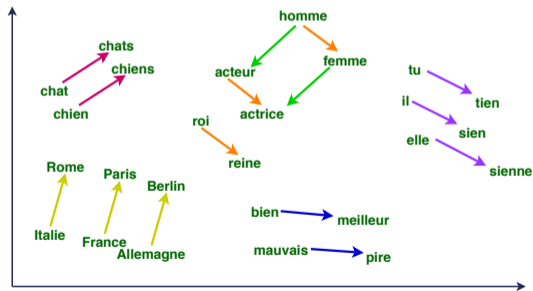
[Mikolov 13]

Métrique fine entre les produits

[Grbovic 15]



Paradigme Word2Vec



⇒ Espace vectoriel sémantique induit

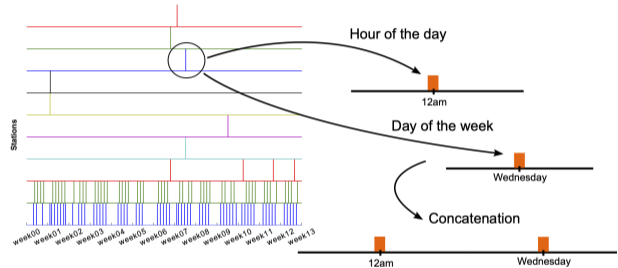
Contribution (1) : Analyse de masses d'informations bruitées et parcellaires

Cartes RFID = smart card = AFC
Collaboration IdF Mobilités (ex STIF)

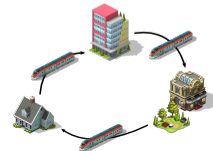


- + Nouvelle vision des transports
- + Evaluation des politiques publiques
- + Aide à la décision
- Données très bruitées
- Entrées seulement

Données disponibles, par usager :



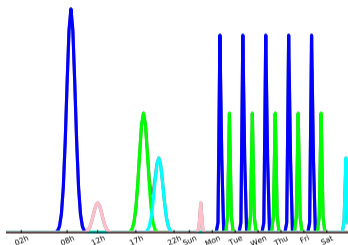
Objectif : séparer / comprendre les usages



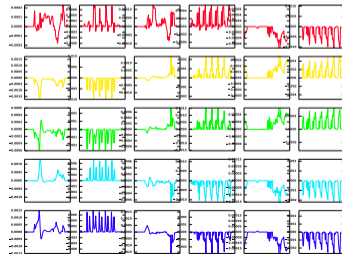
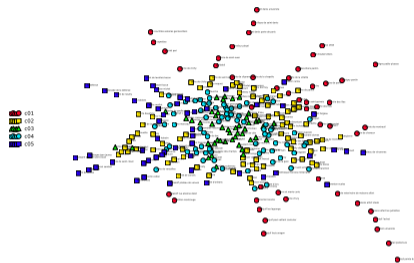
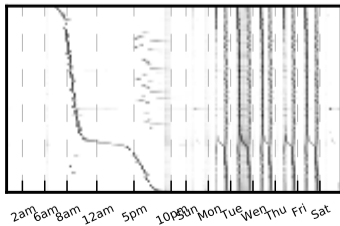
Contribution (1) : Vue générale des résultats obtenus

[Poussevin 14,16]

- 1 Extraction des comportements récurrents individuels
- 2 Différentes échelles fréquentielles (fréquent/rare)
- 3 Projection des comportements sur une carte

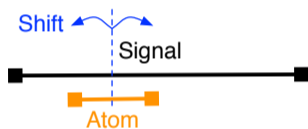


$b = high$



Contribution (1) : Relaxation de la position des motifs caractéristiques

[Tonnelier 16,17,18] NMF invariante en translation : localiser un comportement



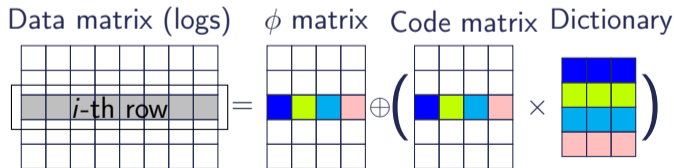
u : usager

w_{uz} : code, puissance du comportement z pour u

ϕ_{uz} : localisation du comportement z pour u

d_z : comportement

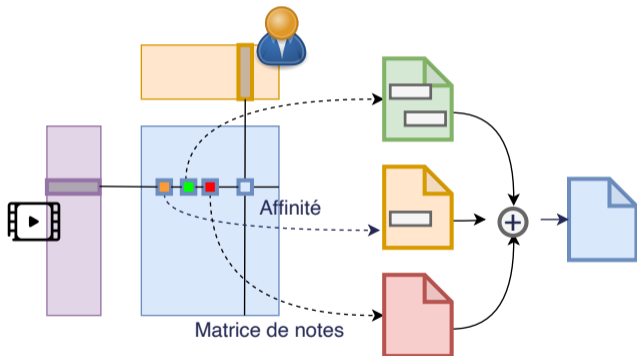
$$u = \sum_z \tau_{u,z} (w_{u,z} d_z) = w_{u,z} d_z (t + \phi_{u,z})$$



Prédiction de comportements / détection d'anomalies / Fact. tensorielle

Contribution (2) : Explique le filtrage collaboratif par données textuelles

[Poussevin 14,15] Combiner prédiction d'affinités et génération de résumé personnalisé



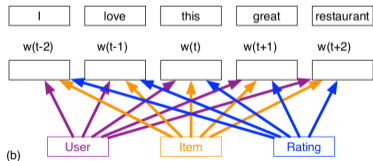
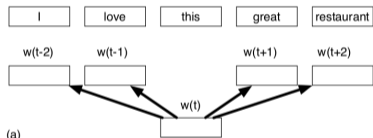
- Mélange d'analyse de sentiments et de recommandation
- Détection des phénomènes de *surprises*
- Difficulté pour l'évaluation quantitative
- Résultats qualitatifs amusants

Un premier pas vers l'xAI

Contribution (3) : Projeter le filtrage collaboratif dans un espace textuel

[Dias 16,17] Réconciliation entre filtrage collaboratif & content based

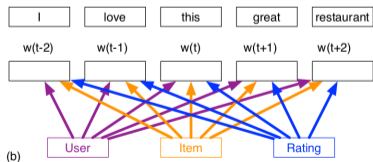
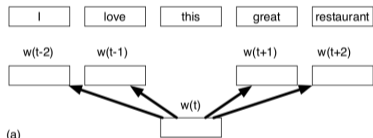
Extension de Word2Vec
+ profils utilisateurs/produits



Contribution (3) : Projeter le filtrage collaboratif dans un espace textuel

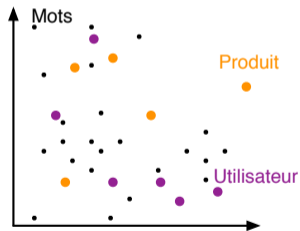
[Dias 16,17] Réconciliation entre filtrage collaboratif & content based

Extension de Word2Vec
+ profils utilisateurs/produits



Espace de représentation multi-modal :

Espace vectoriel unifié



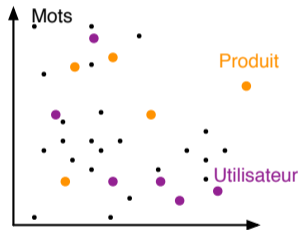
- Prédiction de notes/mots/revues en plus-proches-voisins
- Performances quantitative & qualitative ++

Contribution (3) : Projeter le filtrage collaboratif dans un espace textuel

[Dias 16,17] Réconciliation entre filtrage collaboratif & content based

Espace de représentation multi-modal :

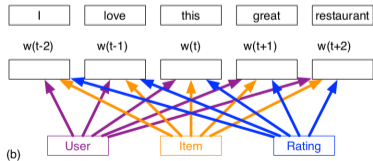
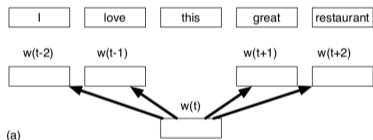
Espace vectoriel unifié



- Prédiction de notes/mots/revues en plus-proches-voisins
- Performances quantitative & qualitative ++

Extension pour le cold-start

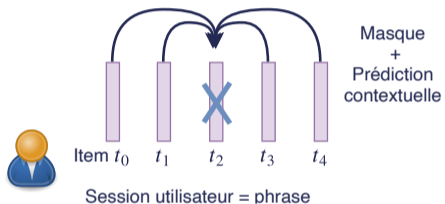
- 1 Construction de profil utilisateur à partir de textes
- 2 Prédiction de notes dans l'espace hybride

Extension de Word2Vec
+ profils utilisateurs/produits

Contribution (4) : Recommandation et dynamique locale

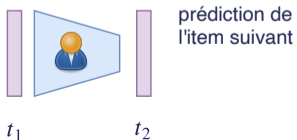
[Guardia-Sebaoun 15,16] : Word2Vec sur des séquences de produits

1. Représentation de produits :



[Chen 12, Grbovic 15]

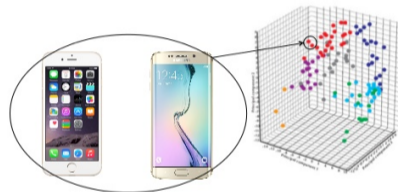
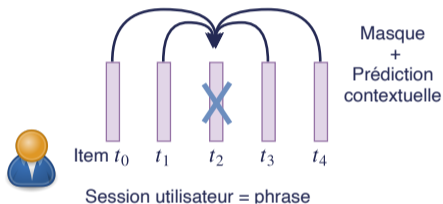
2. Représentation d'utilisateurs :



Contribution (4) : Recommandation et dynamique locale

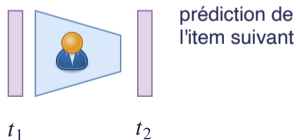
[Guardia-Sebaoun 15,16] : Word2Vec sur des séquences de produits

1. Représentation de produits :



Similarité produit \gg facto. matricielle
(idem Word2Vec \gg pLSA)

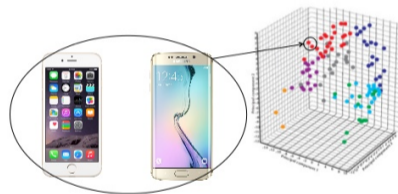
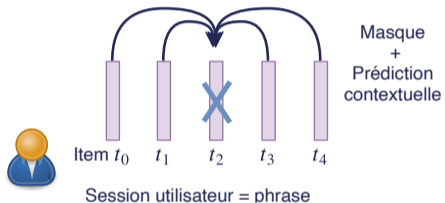
2. Représentation d'utilisateurs :



Contribution (4) : Recommandation et dynamique locale

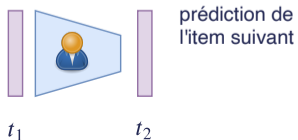
[Guardia-Sebaoun 15,16] : Word2Vec sur des séquences de produits

1. Représentation de produits :



[Chen 12, Grbovic 15]

2. Représentation d'utilisateurs :



Similarité produit \gg facto. matricielle
(idem Word2Vec \gg pLSA)

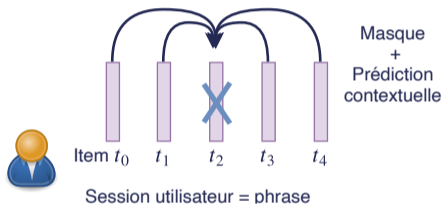
Nouvelle tâche :

Prédiction de la prochaine visite

Contribution (4) : Recommandation et dynamique locale

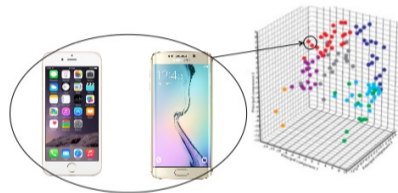
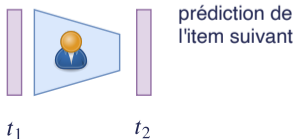
[Guardia-Sebaoun 15,16] : Word2Vec sur des séquences de produits

1. Représentation de produits :



[Chen 12, Grbovic 15]

2. Représentation d'utilisateurs :



Similarité produit \gg facto. matricielle
(idem Word2Vec \gg pLSA)

Nouvelle tâche :

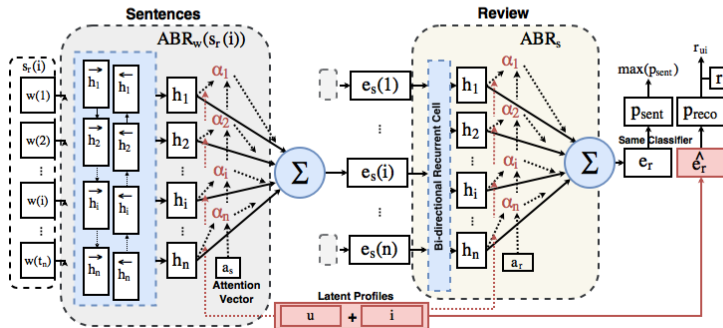
Prédiction de la prochaine visite

Extension projet AMMICO : modélisation
des **visiteurs dans les musées**

Contribution (5) : Vers une sémantique contextuelle

[Dias 18] : RNN + TAL \Rightarrow \nearrow perf. au *niveau phrase*

- Agrégation Mots \Rightarrow Phrases \Rightarrow Documents
- Attention = extraire les éléments clés dans la décision
- Profils reco. pour implémenter l'attention

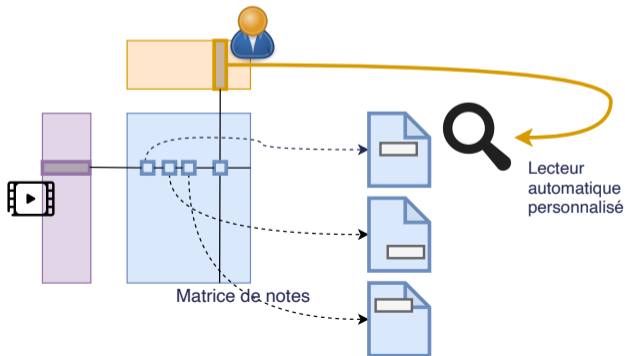


- Sélection personnalisée de mots, phrases, revues \Rightarrow Explications
- \nearrow perf. en recommandation (prédiction de notes)

Contribution (5) : Vers une sémantique contextuelle

[Dias 18] : RNN + TAL \Rightarrow \nearrow perf. au *niveau phrase*

- Agrégation Mots \Rightarrow Phrases \Rightarrow Documents
- Attention = extraire les éléments clés dans la décision
- Profils reco. pour implémenter l'attention



- Sélection personnalisée de mots, phrases, revues \Rightarrow Explications
- \nearrow perf. en recommandation (prédiction de notes)

Conclusions

Focalisation sur les modèles pour expliquer la/les définition(s) de la sémantique.
Les résultats quantitatifs sont dans le mémoire.

- La **sémantique**, ce n'est pas que du texte...
⇒ Encore mieux en **combinant les modalités** !

Conclusions

Focalisation sur les modèles pour expliquer la/les définition(s) de la sémantique.
Les résultats quantitatifs sont dans le mémoire.

- La **sémantique**, ce n'est pas que du texte...
⇒ Encore mieux en **combinant les modalités** !
- *Produits, Mots... Usagers, Utilisateurs* = **mêmes algorithmes** !
- **Ethique** : un faux débat sémantique sur les **modèles**, un vrai débat sur les **usages**

Focalisation sur les modèles pour expliquer la/les définition(s) de la sémantique.
Les résultats quantitatifs sont dans le mémoire.

- La **sémantique**, ce n'est pas que du texte...
⇒ Encore mieux en **combinant les modalités** !
- *Produits, Mots... Usagers, Utilisateurs* = **mêmes algorithmes** !
- **Ethique** : un faux débat sémantique sur les **modèles**, un vrai débat sur les **usages**
- Meilleure prédiction de notes = factorisation matricielle (\neq deep learning)
deep-learning = opportunité de combiner les modalités de données

Architectures de bout en bout et démêlage

Architecture de bout en bout

Retour sur les approches non-supervisées / auto-supervisées

- 1 Encodage de données de plus en plus complexes
- 2 Raffinage avec d'autres tâches
 - (Denoising) Auto-encodeur vectoriel
 - Word2Vec
 - Sequence 2 sequence
 - Transformer / BERT

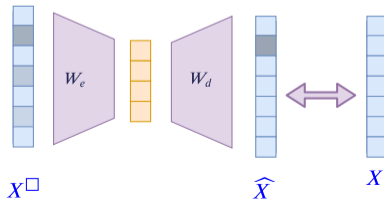


[Erhan 09]

[Mikolov 13]

[Sutskever 14]

[Waswani 17, Devlin 18]

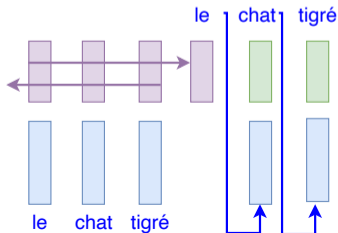


Architecture de bout en bout

Retour sur les approches non-supervisées / auto-supervisées

- 1 Encodage de données de plus en plus complexes
- 2 Raffinage avec d'autres tâches

- (Denoising) Auto-encodeur vectoriel
- Word2Vec
- Sequence 2 sequence
- Transformer / BERT



[Erhan 09]

[Mikolov 13]

[Sutskever 14]

[Waswani 17, Devlin 18]

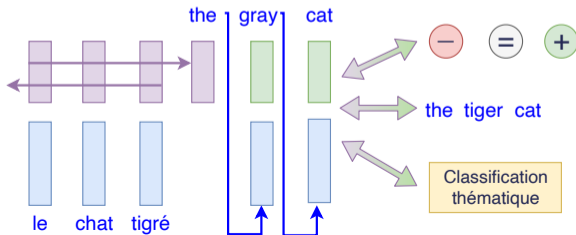
- Appris sur de large corpus

Architecture de bout en bout

Retour sur les approches non-supervisées / auto-supervisées

- 1 Encodage de données de plus en plus complexes
- 2 Raffinage avec d'autres tâches

- (Denoising) Auto-encodeur vectoriel
- Word2Vec
- Sequence 2 sequence
- Transformer / BERT



[Erhan 09]

[Mikolov 13]

[Sutskever 14]

[Waswani 17, Devlin 18]

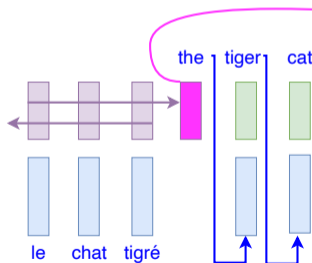
- Appris sur de large corpus
- Transféré sur d'autres applications
Encoder et/ou Decoder
- Raffiné / modifié / étendu
avec d'autres données

Architecture de bout en bout & démêlage / *disentanglement*

Interpréter l'espace de représentation

- Variational AutoEncoder
- Generative Adversarial Network
- Encodage / Transformation / Génération

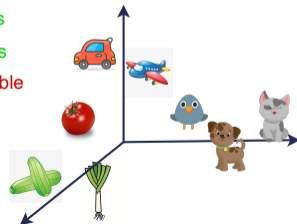
[Kingma 14]
[Goodfellow 14, Chen 16]
[Lample 19]



Représentation intermédiaire

- + Reconstruction des données
- + Encodage de connaissances
- Espace latent non interprétable

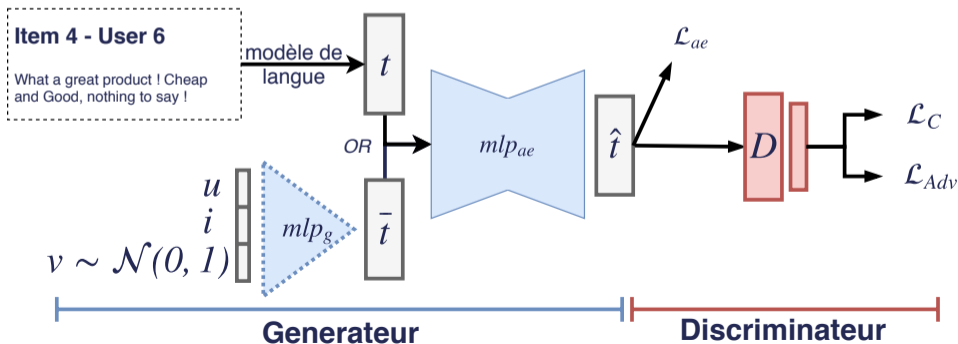
Objectif:



Contribution (1) : Modèle génératif adversaire entre textes et profils

[Dias 19] La recommandation comme une prédiction de sentiments en aveugle :
Prédire une note =

- 1 Générer la revue pour l'utilisateur
- 2 Noter cette revue



Contribution (1) : Modèle génératif adversaire entre textes et profils

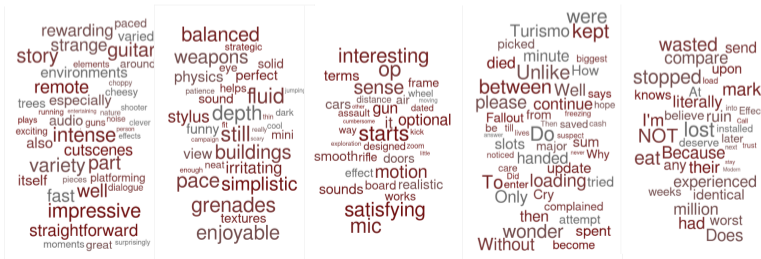
[Dias 19] La recommandation comme une prédiction de sentiments en aveugle :

Prédire une note =

- 1 Générer la revue pour l'utilisateur
- 2 Noter cette revue

- Bonne prédiction notes + textes
- Espace de représentation \Leftrightarrow dimensions explicites en texte

Exemple d'axe latent associé à du texte :



Contribution (2) : Encodeur-décodeur pour la robustesse et la dynamique

[Dias 17] Données bruitées & structurée = CVs & dynamique = prochain job

- Job \neq catégorie (différentes écritures)
- CV = très bruités (titres longs, fautes orthographiques)

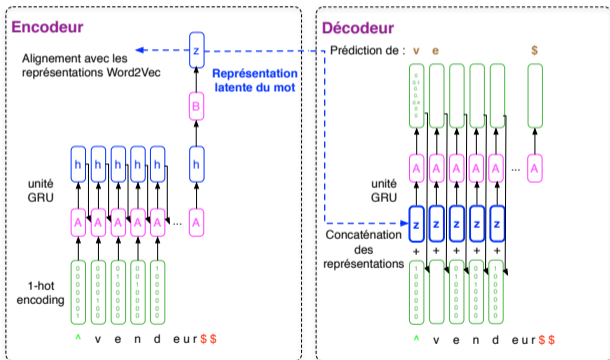
CV de Mirella Mercier

Prédiction (pointe vers le titre 'Pharmacienne')

Encodage (pointe vers le titre 'Responsable de recherche pharmaceutique')

1-hot encoding (pointe vers les matrices de caractères)

1-hot encoding (pointe vers les matrices de caractères)



Archi. char-to-char

Contribution (2) : Encodeur-décodeur pour la robustesse et la dynamique

[Dias 17] Données bruitées & structurée = CVs & dynamique = prochain job

- Job \neq catégorie
(différentes écritures)
- CV = très bruités
(titres longs, fautes orthographiques)



Profil
Mirella Mercier
8 décembre 1968
Lieu de naissance: Marseille

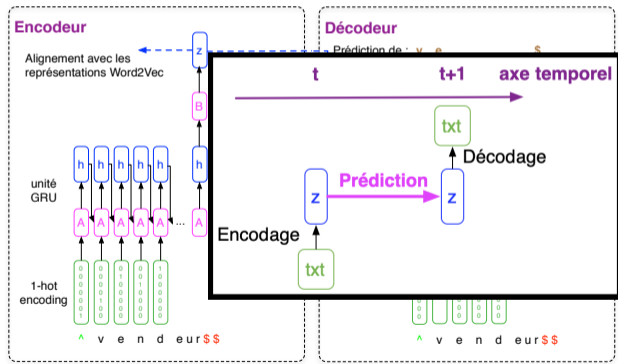
Compétences
Pharmacologie ●●●●●
Micro-biologie ●●●●●
Suite Office ●●●●●
G-Stock Pharma ●●●●●

Langues
Français ●●●●●
Italien ●●●●●
Allemand ●●●●●

12/2010 - présent
Pharmacienne
Pharmacie de l'Arbre du Bois, Lyon
Au sein de cette petite pharmacie, je suis en charge des missions suivantes:
• Gestion administrative (facturation, commandes, etc.)
• Préparation, analyse et vérification d'ordonnances des patients
• Exercice pharmaceutique
• Pharmacovigilance, pharmacodépendance

12/2003 - 08/2010
Responsable de recherche pharmaceutique
Boiron Laboratoires, Lyon
Au sein des laboratoires Boiron, je suis responsable d'un programme de recherche sur les anti-coagulants.

Annotations:
- **Prédiction** (pointe vers 'Pharmacienne')
- **Encodage** (pointe vers 'Responsable de recherche pharmaceutique')

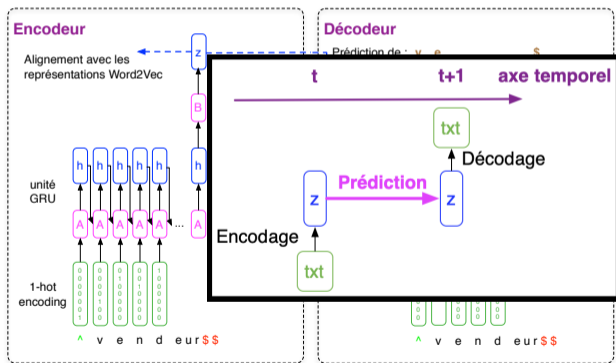


Archi. char-to-char

Contribution (2) : Encodeur-décodeur pour la robustesse et la dynamique

[Dias 17] Données bruitées & structurée = CVs & dynamique = prochain job

- Job \neq catégorie
(différentes écritures)
- CV = très bruités
(titres longs,
fautes orthographiques)
- Aide au remplissage de
formulaire
- Prédiction de churn
- Caractérisation des carrières



Archi. char-to-char

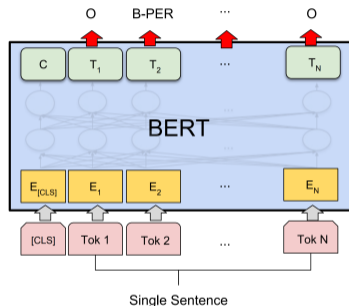
Contribution (3) : De la sémantique aux connaissances

NER : Named Entity Recognition = base de l'extraction de connaissances

When **Sebastian Thrun** PERSON started at **Google** ORG in **2007** DATE, few people outside of the company took him seriously. "I can tell you very senior CEOs of major **American** NORP car companies would shake my hand and turn away because I wasn't worth talking to," said **Thrun** PERSON, now the co-founder and CEO of online higher education startup Udacity, in an interview with **Recode** ORG **earlier this week** DATE.

A little **less than a decade later** DATE, dozens of self-driving startups have cropped up while automakers around the world clamor, wallet in hand, to secure their place in the fast-moving world of fully automated transportation.

- HMM puis CRF
- Bi-LSTM + CRF + char & word encoding
- ELMO - BERT



[Zhou 02, McCallum 03]

[Lample 16]

[Peters 18, Devlin 18]

Contribution (3) : Entités nommées, reconnaissance ou extraction ?

[Taillé 19,20] Problème dans la constitution des jeux de données de référence :

Train

JF Kennedy was fatally shot by former U.S. Marine **Lee Harvey Oswald**

Test

John Fitzgerald Kennedy, the **35th president of the United States**, is assassinated by **Lee Harvey Oswald** while traveling through **Dallas**

Reconnaitre un terme \neq Extraire un terme inconnu

Contribution (3) : Entités nommées, reconnaissance ou extraction ?

[Taillé 19,20] Problème dans la constitution des jeux de données de référence :

EM = Exact Match entre train/test

PM = Partial Match entre train/test

Train

JF Kennedy was fatally shot by former U.S. Marine **Lee Harvey Oswald**

Test

John Fitzgerald Kennedy, the **35th president of the United States**, is assassinated by **Lee Harvey Oswald** while traveling through **Dallas**

Reconnaitre un terme \neq Extraire un terme inconnu

Contribution (3) : Entités nommées, reconnaissance ou extraction ?

[Taillé 19,20] Problème dans la constitution des jeux de données de référence :

EM = Exact Match entre train/test

PM = Partial Match entre train/test

| | | CoNLL03 | | | | | OntoNotes* | | | | | WNUT* | | | |
|-------|-----|---------|-----|-----|-----|------------|------------|-----|-----|-----|------------|-------|-----|-----|------------|
| | | LOC | MIS | ORG | PER | ALL | LOC | MIS | ORG | PER | ALL | LOC | ORG | PER | ALL |
| Self | EM | 82% | 67% | 54% | 14% | 52% | 87% | 93% | 54% | 49% | 69% | - | - | - | - |
| | PM | 4% | 11% | 17% | 43% | 20% | 6% | 2% | 32% | 36% | 20% | 11% | 5% | 13% | 12% |
| | New | 14% | 22% | 29% | 43% | 28% | 7% | 5% | 14% | 15% | 11% | 89% | 95% | 87% | 88% |
| CoNLL | EM | - | - | - | - | - | 70% | 78% | 18% | 16% | 42% | 26% | 8% | 1% | 7% |
| | PM | - | - | - | - | - | 7% | 10% | 45% | 46% | 28% | 9% | 15% | 16% | 14% |
| | New | - | - | - | - | - | 23% | 12% | 38% | 38% | 30% | 65% | 77% | 83% | 78% |

Reconnaitre un terme \neq Extraire un terme inconnu

Contribution (3) : entités nommées, reconnaissance ou extraction ?

[Taillé 19,20]

Modèles de langue récents = principaux vecteurs d'amélioration en NER

| Embedding | Dim | CoNLL03 | | | | OntoNotes* | | | | WNUT* | | |
|--------------|------|--------------------|--------------------|--------------------|---------------------------|--------------------|--------------------|--------------------|---------------------------|---------------------|---------------------|----------------------------|
| | | EM | PM | New | All | EM | PM | New | All | PM | New | All |
| BERT | 4096 | 95.7 _{.1} | 88.8 _{.3} | 82.2 _{.3} | 90.5 _{.1} | 96.9 _{.2} | 88.6 _{.3} | 81.1 _{.5} | 93.5 _{.2} | 77.0 _{4.6} | 53.9 _{.9} | 57.0 _{1.0} |
| ELMo | 1024 | 95.9 _{.1} | 89.2 _{.5} | 85.8 _{.7} | 91.8 _{.3} | 97.1 _{.2} | 88.0 _{.2} | 79.9 _{.7} | 93.4 _{.2} | 67.7 _{3.2} | 49.5 _{.9} | 52.1 _{1.0} |
| Flair | 4096 | 95.4 _{.1} | 88.1 _{.6} | 83.5 _{.5} | 90.6 _{.2} | 96.7 _{.1} | 85.8 _{.5} | 75.0 _{.6} | 92.1 _{.2} | 64.9 _{.7} | 48.2 _{2.0} | 50.4 _{1.8} |
| ELMo[0] | 1024 | 95.8 _{.1} | 87.2 _{.2} | 83.5 _{.4} | 90.7 _{.1} | 96.9 _{.1} | 85.9 _{.3} | 75.5 _{.6} | 92.4 _{.1} | 72.8 _{1.3} | 45.4 _{2.8} | 49.1 _{2.3} |
| GloVe + char | 350 | 95.3 _{.3} | 85.5 _{.7} | 83.1 _{.7} | 89.9 _{.5} | 96.3 _{.1} | 83.3 _{.2} | 69.9 _{.6} | 91.0 _{.1} | 63.2 _{4.6} | 33.4 _{1.5} | 38.0 _{1.7} |
| GloVe | 300 | 95.1 _{.4} | 85.3 _{.5} | 81.1 _{.5} | 89.3 _{.4} | 96.2 _{.2} | 82.9 _{.2} | 63.8 _{.5} | 90.4 _{.2} | 59.1 _{2.9} | 28.1 _{1.5} | 32.9 _{1.2} |

Contribution (3) : entités nommées, reconnaissance ou extraction ?

[Taillé 19,20]

Modèles de langue récents = principaux vecteurs d'amélioration en NER

| Embedding | Dim | CoNLL03 | | | | OntoNotes* | | | | WNUT* | | |
|--------------|------|--------------------|--------------------|--------------------|--------------------------|--------------------|--------------------|--------------------|--------------------------|---------------------|---------------------|---------------------------|
| | | EM | PM | New | All | EM | PM | New | All | PM | New | All |
| BERT | 4096 | 95.7 _{.1} | 88.8 _{.3} | 82.2 _{.3} | 90.5 _{.1} | 96.9 _{.2} | 88.6 _{.3} | 81.1 _{.5} | 93.5_{.2} | 77.0 _{4.6} | 53.9 _{.9} | 57.0_{1.0} |
| ELMo | 1024 | 95.9 _{.1} | 89.2 _{.5} | 85.8 _{.7} | 91.8_{.3} | 97.1 _{.2} | 88.0 _{.2} | 79.9 _{.7} | 93.4_{.2} | 67.7 _{3.2} | 49.5 _{.9} | 52.1 _{1.0} |
| Flair | 4096 | 95.4 _{.1} | 88.1 _{.6} | 83.5 _{.5} | 90.6 _{.2} | 96.7 _{.1} | 85.8 _{.5} | 75.0 _{.6} | 92.1 _{.2} | 64.9 _{.7} | 48.2 _{2.0} | 50.4 _{1.8} |
| ELMo[0] | 1024 | 95.8 _{.1} | 87.2 _{.2} | 83.5 _{.4} | 90.7 _{.1} | 96.9 _{.1} | 85.9 _{.3} | 75.5 _{.6} | 92.4 _{.1} | 72.8 _{1.3} | 45.4 _{2.8} | 49.1 _{2.3} |
| GloVe + char | 350 | 95.3 _{.3} | 85.5 _{.7} | 83.1 _{.7} | 89.9 _{.5} | 96.3 _{.1} | 83.3 _{.2} | 69.9 _{.6} | 91.0 _{.1} | 63.2 _{4.6} | 33.4 _{1.5} | 38.0 _{1.7} |
| GloVe | 300 | 95.1 _{.4} | 85.3 _{.5} | 81.1 _{.5} | 89.3 _{.4} | 96.2 _{.2} | 82.9 _{.2} | 63.8 _{.5} | 90.4 _{.2} | 59.1 _{2.9} | 28.1 _{1.5} | 32.9 _{1.2} |

Modèle de langue = état de l'art sur les benchmarks de référence

Contribution (3) : entités nommées, reconnaissance ou extraction ?

[Taillé 19,20]

Modèles de langue récents = principaux vecteurs d'amélioration en NER

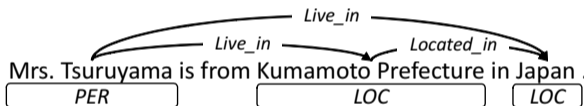
| Embedding | Dim | CoNLL03 | | | | OntoNotes* | | | | WNUT* | | |
|--------------|------|--------------------|--------------------|--------------------|--------------------------|--------------------|--------------------|--------------------|--------------------------|---------------------|---------------------|---------------------------|
| | | EM | PM | New | All | EM | PM | New | All | PM | New | All |
| BERT | 4096 | 95.7 _{.1} | 88.8 _{.3} | 82.2 _{.3} | 90.5 _{.1} | 96.9 _{.2} | 88.6 _{.3} | 81.1 _{.5} | 93.5_{.2} | 77.0 _{4.6} | 53.9 _{.9} | 57.0_{1.0} |
| ELMo | 1024 | 95.9 _{.1} | 89.2 _{.5} | 85.8 _{.7} | 91.8_{.3} | 97.1 _{.2} | 88.0 _{.2} | 79.9 _{.7} | 93.4_{.2} | 67.7 _{3.2} | 49.5 _{.9} | 52.1 _{1.0} |
| Flair | 4096 | 95.4 _{.1} | 88.1 _{.6} | 83.5 _{.5} | 90.6 _{.2} | 96.7 _{.1} | 85.8 _{.5} | 75.0 _{.6} | 92.1 _{.2} | 64.9 _{.7} | 48.2 _{2.0} | 50.4 _{1.8} |
| ELMo[0] | 1024 | 95.8 _{.1} | 87.2 _{.2} | 83.5 _{.4} | 90.7 _{.1} | 96.9 _{.1} | 85.9 _{.3} | 75.5 _{.6} | 92.4 _{.1} | 72.8 _{1.3} | 45.4 _{2.8} | 49.1 _{2.3} |
| GloVe + char | 350 | 95.3 _{.3} | 85.5 _{.7} | 83.1 _{.7} | 89.9 _{.5} | 96.3 _{.1} | 83.3 _{.2} | 69.9 _{.6} | 91.0 _{.1} | 63.2 _{4.6} | 33.4 _{1.5} | 38.0 _{1.7} |
| GloVe | 300 | 95.1 _{.4} | 85.3 _{.5} | 81.1 _{.5} | 89.3 _{.4} | 96.2 _{.2} | 82.9 _{.2} | 63.8 _{.5} | 90.4 _{.2} | 59.1 _{2.9} | 28.1 _{1.5} | 32.9 _{1.2} |

Le phénomène est encore bien plus fort dans les cas difficiles !

Contribution (4) Performances en extraction de relation de bout en bout

[Taillé 20] : Multiplication des métriques \Rightarrow confusion + *optimisation*

- Types des entités
- Bornes de entités
- Types des relations

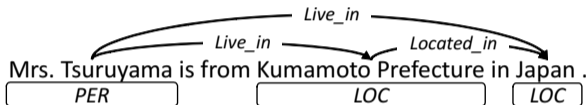


Analyse des modèles de l'état de l'art dans différentes situations (métriques / ablations)

Contribution (4) Performances en extraction de relation de bout en bout

[Taillé 20] : Multiplication des métriques \Rightarrow confusion + *optimisation*

- Types des entités
- Bornes de entités
- Types des relations



Analyse des modèles de l'état de l'art dans différentes situations (métriques / ablations)

Problème de définition de la tâche :

Entités **non** vues
↓
Knowledge **extraction**

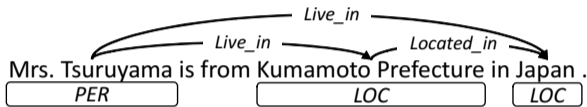
Analyse de la phrase :

Comprendre la position des entités
et la nature de la relation

Contribution (4) Performances en extraction de relation de bout en bout

[Taillé 20] : Multiplication des métriques \Rightarrow confusion + *optimisation*

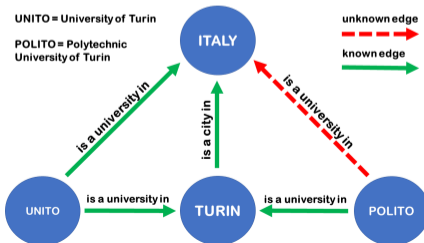
- Types des entités
- Bornes de entités
- Types des relations



Analyse des modèles de l'état de l'art dans différentes situations (métriques / ablations)

Problème de définition de la tâche :

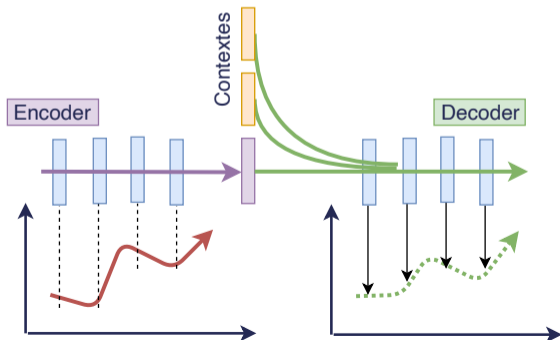
Entités **déjà vues**
 \Downarrow
 Knowledge **completion**



Contribution (5) Génération de séries temporelles contextualisées

[Cribier-Delande 19,20] Modèles génératifs pour l'analyse de séries temporelles

- Modèles AR [Box 68] ⇒ Inapte à la prédiction à long terme
- + Saisonnalité ⇒ Robuste
- Modèles basés sur la saisonnalité + ML [Taylor 17] ⇒ Très efficace + robuste



Contexte = super saisonnalité

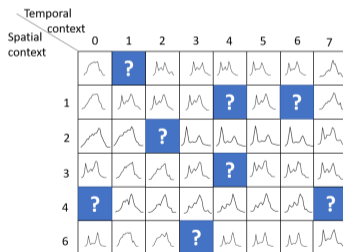
- Jour / mois / saison
- Météo
- Géographie
- Événements...

Contribution (5) Génération de séries temporelles contextualisées

[Cribier-Delande 19,20] Modèles génératifs pour l'analyse de séries temporelles

- Modèles AR [Box 68] ⇒ Inapte à la prédiction à long terme
- + Saisonnalité ⇒ Robuste
- Modèles basés sur la saisonnalité + ML [Taylor 17] ⇒ Très efficace + robuste

Nouvelle
formulation
prédictive :



Préd. affluence transports en commun :
contexte = jour/station

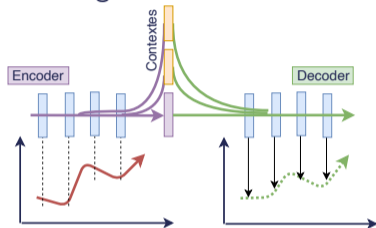
Prédiction à court terme / long terme ⇒ **Modèle efficace partout !**

Contexte = super saisonnalité

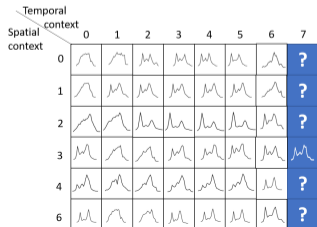
- Jour / mois / saison
- Météo
- Géographie
- Événements...

Contribution (5) Génération de séries temporelles contextualisées

Encodage du contexte :

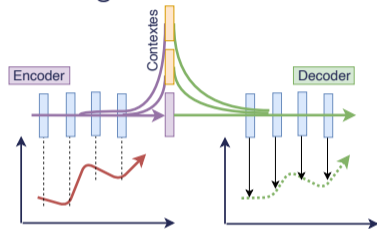


Nouveau type de problème :

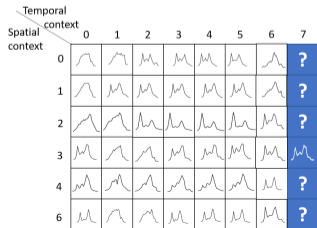


Contribution (5) Génération de séries temporelles contextualisées

Encodage du contexte :

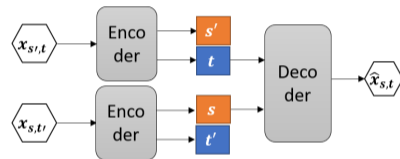


Nouveau type de problème :



[Cribier-Delande 20]

Assurer le démêlage des facteurs :



- Projections en affluence transport : nouvelle station / nouveau contexte
- Prédiction de la conso. électrique
- Prédiction de la pollution

Performances de l'état de l'art dans tous les domaines applicatifs

Conclusion & Projet de recherche

Résumé des contributions

Sémantique multi-modale :

- fMRI & données textuelles [Pipanmaekaporn 14,15]
- Profils & données textuelles [Poussevin 14,15 Dias 16,17,18,19 Gainon 20]
- Profils & dynamique temporelle [Guardia 15,16 Dias 17]
- Profils & séries temporelles [Cribier-Delande 19,20]

De la sémantique aux connaissances :

- Entités nommées & relations [Simon 19 Taille 19,20]

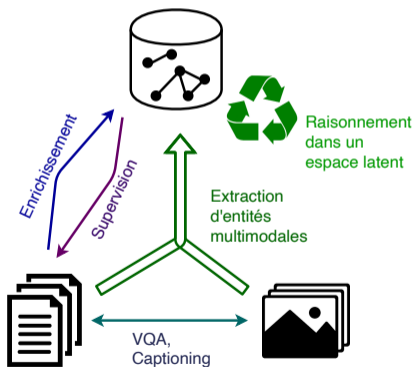
Au delà des données iid, la sémantique

retrouver une partie manquante des données / travailler dans un domaine connexe /
transposer le contexte

⇒ conserver les performances du modèle

Perspectives (1) : vers des entités multi-modales

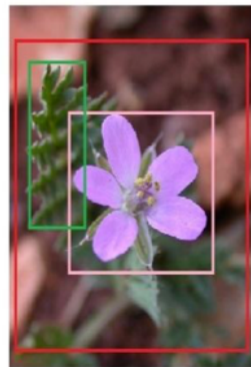
Collaboration avec le Museum National d'Histoire Naturelle (R. VignesLebbe)
Doctorat de Maya Sahraoui (Décembre 2020)



Erodium romanum Wild :

Plante vivace, plus ou moins poilue, acaule, de 10-30 cm. Souche courte, épaisse, non écailleuse. Feuilles toutes radicales, longues, pennatiséquées, à segments écartés, larges, ovales, incisés-pennatifides. Fleurs rouges, assez grandes, 4-8 en ombelles sur des pédoncules radicaux ordinairement plus courts que les feuilles

Base de connaissances :
Aide à la décision (supervision faible)



Perspective (2) : vers un modèles de langue pour les séries temporelles

Collaboration EDF (G. Agoua) & LIP6 (N. Baskiotis)
Doctorat de **Etienne Le Naour** (S1 2021)

- Modèles **images** transférables depuis [Krizhevsky 12]
- Modèles **textes** transférables (+ multilingues) depuis [Collobert 08]
- Les séries temporelles posent problèmes :
 - trop de classes différentes
 - contextes applicatifs très différents
 - trop de bruits

Mais les séries temporelles sont adaptés à l'auto-apprentissage

⇒ Extraction **non supervisée** des contextes / distinction des classes de séries temporelles

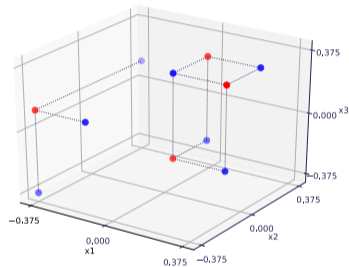
⇒ Construction de **modèles auto-appris** transférables dans une classe de problèmes

Perspective (3) : vers des algorithmes de profiling plus transparents

Collaboration Deezer (R. Hennequin)

Doctorat de **Darius Afchar** (Décembre 2020)

- Feature Attribution au niveau des instances
- Explication des suggestions
- Structuration, supervisée ou non, de l'espace de représentation



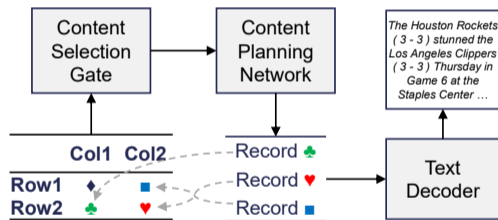
Un profil de recommandation :

- (1) expliquant sur quelles données il se base pour décider/suggérer
- (2) interrogeable par l'utilisateur
- (3) modifiable par l'utilisateur en retour

Perspective (4) : génération de textes à partir de données

Collaboration LIP6 (L. Soulier)
Doctorat de **Tristan Luiggi** (Mars 2021)

- Extraction d'informations à partir de textes / tableaux / données brutes
- Génération de textes résumés
- Contrôle de la qualité des textes générés



Concevoir un algorithme capable d'expliquer ce qu'il est en train de faire

xIA & Acceptabilité des algorithmes

l'IA transforme la société...

Cette transformation doit être une amélioration !

Au niveau politique

- Contrôle des usages
(reconnaissance faciale)
- Contrôle de la collecte des données
(RGPD)
- Limite sur le droit de manipuler des données
(statistiques ethniques)

Au niveau scientifique

- Former des spécialistes responsables
- Informer les non-spécialistes
- Concevoir des algorithmes plus transparents
- Concevoir des algorithmes plus interactifs avec l'utilisateur