



# REDS MÉTHODOLOGIE RECHERCHE

## Modèles

Tuesday 12<sup>th</sup> January, 2021

Laure Soulier



# Topologie des contributions scientifiques

# Topologie des contributions scientifiques

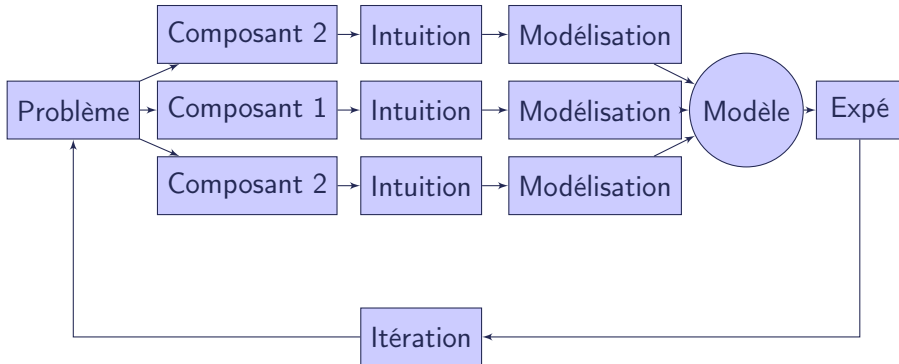
- **Proposition de modèle**
  - Modèle de recommandation
  - Modèle de prédiction
  - Modèle de recherche d'information
  - ...
- **Analyse empirique**
  - Comparaison des différents modèles de recommandation
  - Analyse comparative des architectures neuronales pour une problématique de classification
  - Analyse de l'impact de l'âge sur le comportement de recherche des utilisateurs
- **Prototypage**
  - Moteurs de recherche
  - Interface de prédiction
- **Jeux de données**
  - Données labellisées (classification, nettoyage, etc...)
  - Logs de comportements utilisateurs
- **Synthèse de l'état de l'art (généralement en fin de thèse)**

Processus de "création" d'un  
modèle

# Processus de "création" d'un modèle

## Définition

Modèle : Formalisation mathématique qui répond à une problématique



Et en pratique...

## En pratique...

- Modification de la fonction de coût
  - Exemple :

by minimizing the classification loss on the training data:

$$\sum_{i \in \mathcal{T}} \Delta(f_{\theta}(z_i), y_i) \quad (3)$$

where  $\Delta(f_{\theta}(z_i), y_i)$  is the loss to predict  $f_{\theta}(z_i)$  instead of the real label  $y_i$ , and  $\mathcal{T}$  is the training set.

In order to make sure that connected nodes have similar representations, we introduce the other following loss:

$$\sum_{i,j:w_{i,j} \neq 0} w_{i,j} \|z_i - z_j\|^2 \quad (4)$$

which forces the approach of the latent representation of connected nodes. The complete loss function is the aggregation of the classification and similarity loss:

$$L(z, \theta) = \sum_{i \in \mathcal{T}} \Delta(f_{\theta}(z_i), y_i) + \lambda \sum_{i,j:w_{i,j} \neq 0} w_{i,j} \|z_i - z_j\|^2 \quad (5)$$

Figure 1: [Nozza et al., ICSEFM 2014]

# En pratique...

- Pré-traitement/Encodage des données
  - Exemple : Extraction de relations dans un texte avec des réseaux de neurones

|                      |   |
|----------------------|---|
| Texte brut           | → sparsité  |
| Word embedding + POS | → prise en compte de la sémantique et de la grammaire |

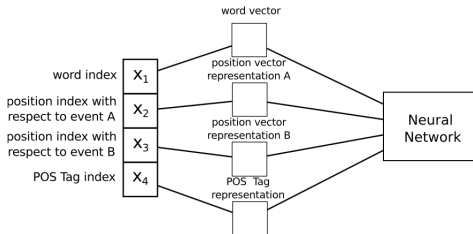


Figure 2: [Gerald, 2016]



# En pratique...

- Modification d'une architecture
  - Prise en compte de la "connaissace globale" issue des textes dans un LSTM

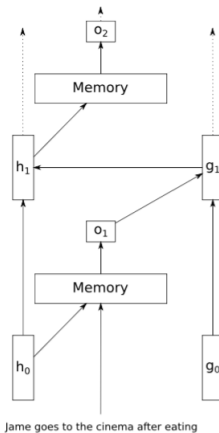


Figure 3: [Gerald, 2016]

# En pratique...

- Transfert de méthodologie d'un domaine
  - Exploiter les réseaux de neurones en RI

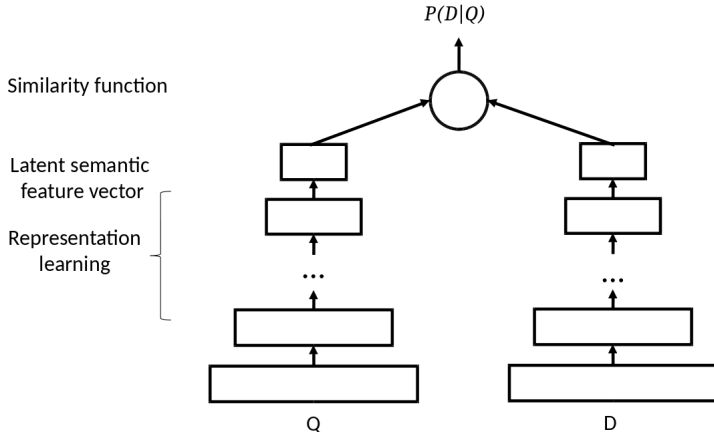


Figure 4: [Nguyen et al., 2017]