



CentraleSupélec

Deep Unordered Composition Rivals Syntactic Methods for Text Classification

Mohit Iyyer, Varun Manjunatha, Jordan Boyd-Graber, Hal
Daumé III

Université du Maryland, Université du Colorado

Oscar Pastural, Louis Gauthier, Clement Florval

January 6, 2025

Contexte : Introduction à la NLP en 2015

- ▶ **Conférence** : 53ème réunion de l'ACL et 7ème conférence sur le NLP à Pékin (2015).
- ▶ **Problématique** : Représenter le contenu textuel pour des tâches comme la classification.
- ▶ **Tendance** : Recherche de modèles plus simples, rapides et performants pour remplacer les solutions syntaxiques complexes.

Avancées et méthodes dominantes en 2015

- ▶ **Méthodes syntaxiques dominantes :**
 - ▶ Recursive Neural Networks (RecNN), TreeLSTM, CNN.
 - ▶ Utilisation de parse trees pour capturer les relations syntaxiques.
- ▶ **Inconvénients majeurs :**
 - ▶ Coût computationnel élevé.
 - ▶ Dépendance à une structure syntaxique consistante.
- ▶ **Avancées récentes :**
 - ▶ Premiers succès de l'apprentissage profond grâce aux embeddings préentraînés (e.g., GloVe).

Présentation de la méthode : Deep Averaging Network (DAN)

Étapes principales du DAN :

1. Moyenne des vecteurs d'embedding des mots (**NBOW**).
2. Transformation par des couches non linéaires (ReLU/tanh).
3. Classification via une couche softmax finale.
4. Régularisation par **word dropout**, supprimant aléatoirement des mots pendant l'entraînement.

Points clés :

- ▶ Approche non syntaxique, mais profondeur compensant l'absence d'ordre.
- ▶ Simple, rapide et efficace même sur des ressources limitées.

Comparaison aux modèles syntaxiques

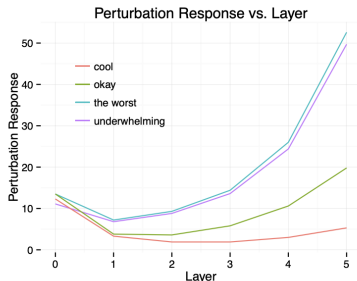
Modèles dominants en 2015 :

- ▶ **Recursive Neural Networks (RecNN)** :
 - ▶ Dépendent des parse trees, coûteux en calcul.
 - ▶ Performants sur des tâches nécessitant la syntaxe (e.g., négation).
- ▶ **TreeLSTM et RecNTN** :
 - ▶ Extensions plus puissantes mais encore plus lourdes computationnellement.
- ▶ **CNN** :
 - ▶ Capturent les relations locales, mais temps d'entraînement élevé.

Avantages du DAN :

- ▶ **Simplicité** : Aucun besoin de parse trees.
- ▶ **Rapidité** : Entraînement possible sur un laptop standard.
- ▶ **Robustesse** : Performances stables même avec des données variées ou bruitées.

Forces du DAN : sensibilité aux variations lexicales



Idée clé :

En augmentant la profondeur (DAN), on capte mieux les changements de mots de polarité différente, contrairement au NBOW superficiel.

En pratique :

Un mot pivot comme "awesome" remplacé par "underwhelming" ou "the worst" a un impact de plus en plus marqué à mesure que le réseau s'approfondit.

Faiblesses du DAN : difficulté à capturer la négation

Problème :

Sans structure syntaxique ni récurrence, le DAN peine à modéliser des phénomènes complexes tels que la double négation.

Note :

Même les modèles syntaxiques de l'époque (RecNN, DRecNN) avaient eux aussi du mal avec la négation. Le DAN se situe donc dans un contexte global de difficulté à traiter ce phénomène.

Sentence	DAN	DRecNN	Ground Truth
a fussy movie that's not merely unwatchable , but also unlistenable	negative	negative	negative
if you're not a prepubescent girl, you'll be laughing at briney spears' movie-starring debut whenever it does it	negative	negative	negative
have you impatiently squinting at your watch			
blessed with immense physical growess he may well be, but ahole is simply not an actor	positive	neutral	negative
who knows what exactly godard is on about in this film, but his words and images do it have to add up to mesmerize you.	positive	positive	positive
it's so good that its relentless , polished wit can withstand not only meep school productions , but even oliver parker's movie adaptation	negative	positive	positive
too bad , but thanks to some lovely comedic moments and several fine performances, it's not a total loss	negative	negative	positive
this movie was not good	negative	negative	negative
this movie was good	positive	positive	positive
this movie was bad	negative	negative	negative
this movie was not bad	negative	negative	positive

Benchmarks et résultats

Tâches évaluées :

- ▶ Analyse de sentiment :
 - ▶ SST (fine-grained et binaire), Rotten Tomatoes, IMDB.
- ▶ Question answering :
 - ▶ Quiz Bowl avec données hétérogènes (e.g., Wikipedia).

Résultats principaux :

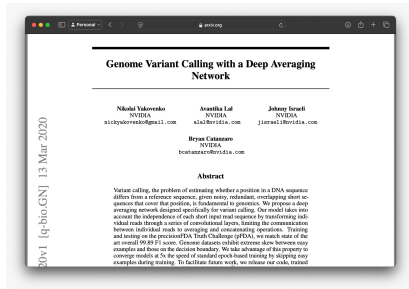
- ▶ Comparable aux modèles syntaxiques complexes :
 - ▶ **SST** : 86,3% vs 86,9% (TreeLSTM).
 - ▶ **IMDB** : Meilleures performances que RecNN et WRRBM.
- ▶ Robustesse avec des données variées :
 - ▶ Meilleure adaptation aux données bruitées (Quiz Bowl avec Wikipedia).
- ▶ Temps d'entraînement réduit : Quelques minutes contre plusieurs heures pour les modèles syntaxiques.

Impact du papier dans le domaine du NLP

- ▶ **Remise en question** : Démonstration qu'une approche simple comme le Deep Averaging Network (DAN) peut rivaliser avec des méthodes syntaxiques complexes pour la classification de texte.
- ▶ **Accessibilité** : Réduction des besoins en ressources computationnelles, rendant le NLP plus accessible.
- ▶ **Avancées clés** :
 - ▶ Introduction et utilité du **word dropout**, optimal à 30%.
 - ▶ Mise en évidence de l'importance des **embeddings préentraînés**.
 - ▶ Capacité à s'adapter à des données hors domaine ou bruitées.
 - ▶ Capture des **nuances subtiles** par des réseaux profonds.

Utilisation et période d'application

- ▶ **Applications diverses :**
 - ▶ Principalement en **classification de texte**.
 - ▶ Utilisé par **NVIDIA** en 2020 pour des applications en **génomique**, telles que le *variant calling*.
- ▶ **Références :** Plus de 1400 citations sur Google Scholar.
- ▶ **Période d'utilisation :** Jusqu'à l'émergence des Transformers (BERT, GPT).
- ▶ **Persistance :** Toujours utile pour des environnements limités en ressources.



Méthodes ayant supplanté le DAN

► Transformers :

- Introduits par le papier *Attention Is All You Need* (2017).
- **BERT** (2018) et **GPT** surpassent le DAN en capturant des **relations contextuelles complexes** grâce aux mécanismes d'attention.

► Établissement de nouveaux standards :

- Préentraînement sur de vastes corpus.
- Adaptation efficace à une large variété de tâches NLP.

Conclusion finale et perspectives

- ▶ **Impact durable** : Le DAN a démontré que des modèles simples peuvent être efficaces, influençant les recherches sur des approches efficaces.
- ▶ **Leçons clés** :
 - ▶ Simplicité vs complexité : Plus de complexité ne signifie pas toujours meilleure performance.
 - ▶ Robustesse face à des données hors domaine ou bruitées.
- ▶ **Supplanté mais inspirant** : Bien que remplacé par des modèles Transformers, le DAN reste un exemple marquant d'innovation dans le NLP.