



**HAL**  
open science

## Oubli Catastrophique et Modèles Neuronaux de Recherche d'Information

Jesus Lovon, Laure Soulier, Karen Pinel-Sauvagnat, Lynda Tamine

► **To cite this version:**

Jesus Lovon, Laure Soulier, Karen Pinel-Sauvagnat, Lynda Tamine. Oubli Catastrophique et Modèles Neuronaux de Recherche d'Information. 17ème conférence francophone en Recherche d'Information et Application (CORIA 2021), ARIA : Association Francophone de Recherche d'Information (RI) et Applications; Equipe MRIM du Laboratoire d'Informatique de Grenoble., Apr 2021, Grenoble (virtuel), France. pp.1-5. hal-03309985

**HAL Id: hal-03309985**

**<https://hal.sorbonne-universite.fr/hal-03309985v1>**

Submitted on 30 Jul 2021

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

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



Distributed under a Creative Commons Attribution 4.0 International License

---

# Oubli Catastrophique et Modèles Neuronaux de Recherche d'Information

Jesús Lovón-Melgarejo\* — Laure Soulier\*\* — Karen Pinel-Sauvagnat\* — Lynda Tamine\*

\* Université Paul Sabatier, IRIT, Toulouse, France

\*\* Sorbonne Université, CNRS, LIP6, F-75005 Paris, France

---

*RÉSUMÉ.* Dans cet article, nous nous plaçons dans le cadre d'une tâche de recherche d'information (RI) adhoc dans un flux de corpus caractérisés par des distributions de données significativement différentes, cas des corpus de différents domaines. Nous étudions précisément dans quelle mesure les modèles neuronaux de RI souffrent du problème de l'oubli catastrophique : l'apprentissage à partir d'un nouveau corpus dans le flux provoque un oubli des connaissances préalablement acquises conduisant à la dégradation des performances du modèle sur les corpus antérieurs du flux. Des expérimentations sur 4 corpus de différents domaines montrent que les 5 modèles d'ordonnancement neuronaux que nous avons évalués souffrent de cet oubli, et qu'une stratégie d'apprentissage tout au long de la vie (lifelong learning) peut atténuer le problème. A travers une approche exploratoire basée sur un modèle de régression, nous avons également identifié les caractéristiques des corpus qui pourraient servir de prédicteurs d'un oubli catastrophique.

*Note:* Cet article est une traduction résumée de l'article (Lovón-Melgarejo et al., 2021) publié à ECIR 2021.

*ABSTRACT.* In this paper, we study in what extent neural ranking models catastrophically forget old knowledge acquired from previously observed domains after acquiring new knowledge (i.e. on new datasets), leading to performance decrease on those domains. Our experiments carried out on 4 datasets and 5 neural IR ranking models show that catastrophic forgetting is prevalent and that a lifelong learning strategy mitigates the problem. Using an explanatory approach built on a regression model, we also identify domain characteristics that can predict catastrophic forgetting.

*Note:* this paper is a summary of the ECIR 2021 paper (Lovón-Melgarejo et al., 2021).

*MOTS-CLÉS :* approches neuronales pour la RI, oubli catastrophique, apprentissage tout au long de la vie

*KEYWORDS:* neural IR, catastrophic forgetting, lifelong learning

---

## 1. Contexte

L’oubli catastrophique (*catastrophic forgetting*) des modèles neuronaux est un problème bien connu en apprentissage automatique : l’entraînement séquentiel d’un modèle sur un nouveau domaine provoque un oubli (ou interférence) important des connaissances préalablement apprises sur les précédents domaines provoquant une dégradation des performances sur ces derniers. Son effet a été particulièrement étudié sur des tâches de classification, notamment en vision artificielle (Kirkpatrick *et al.*, 2016; Li et Hoiem, 2018) et Traitement Automatique de la Langue (TAL) (de Masson d’Autume *et al.*, 2019; Mosbach *et al.*, 2020; Thompson *et al.*, 2019; Wiese *et al.*, 2017). La littérature montre que l’impact de l’oubli catastrophique varie en fonction des approches neuronales et des domaines considérés. Cependant, ce phénomène n’a jamais été clairement étudié en Recherche d’Information (RI). Peu d’enseignements existent à ce jour sur le fait que les modèles neuronaux de RI y sont confrontés, ainsi que sur leur transférabilité entre corpus de différents domaines. Pourtant, comprendre sous quelles conditions les modèles oublient de la connaissance préalablement acquise et si une stratégie d’apprentissage tout au long de la vie (*life-long learning*) (Chen et Liu, n.d.; Parisi *et al.*, 2019) pourrait améliorer leurs performances permettrait d’obtenir de précieuses indications tant dans un point de vue théorique que d’un point de vue expérimental pour de futurs travaux.

## 2. Contribution et résultats principaux

Les travaux que nous avons menés autour de la caractérisation de l’oubli catastrophique des modèles neuronaux en RI sont articulés autour des trois questions de recherche suivantes :

**RQ1:** Observe-t-on de l’oubli catastrophique dans les modèles neuronaux de RI ?

**RQ2:** Si oui, existe-t-il des caractéristiques de corpus qui permettraient de prédire un tel oubli ?

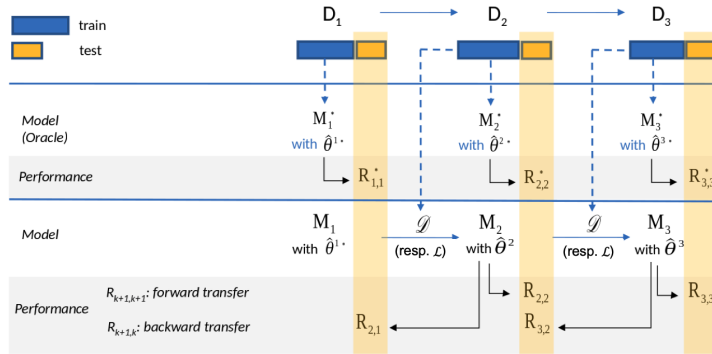
**RQ3:** Est-ce qu’une méthode d’apprentissage tout au long de la vie pourrait atténuer cet oubli ?

### 2.1. Protocole expérimental

Notre protocole expérimental (Cf. figure 1), considère des modèles neuronaux appris et évalués sur un flux de corpus traduisant ainsi un cadre d’apprentissage tout au long de la vie. Les principaux éléments associés sont les suivants :

**Modèles neuronaux pour l’ordonnement des documents :** Nous avons sélectionnés 5 modèles parmi les modèles évalués par (Yang *et al.*, 2019) : 1) modèles basés sur l’interaction : DRMM (Guo *et al.*, 2016), PACRR (Hui *et al.*, 2017) et KNRM (Xiong *et al.*, 2017); 2) modèles basés sur BERT : Vanilla BERT (MacAvaney *et al.*, 2019) and CEDR-KNRM (MacAvaney *et al.*, 2019).

**Corpus :** Nous avons utilisé les 3 corpus suivants : 1) MS MARCO (*ms*) (Bajaj *et al.*, 2016), 2) TREC Microblog (*mb*), 3) TREC CORD19 (*c19*) (Wang *et al.*, 2020).



**Figure 1.** Pipeline expérimental avec une configuration de 3 corpus utilisés en flux pour un modèle  $M$

De nature et de contenu très différents, ces corpus nécessitent l'utilisation de techniques d'apprentissage par transfert (Pan et Yang, 2010). Nous avons considéré 4 configurations utilisant 2 ou 3 corpus (Voir Tableau 1). Comme indiqué dans la littérature (Asghar *et al.*, 2020; Li et Hoiem, 2018), les flux ( $D_1 \rightarrow D_2$ ) ou ( $D_1 \rightarrow D_2 \rightarrow D_3$ ) considèrent les corpus  $D_i$  par ordre décroissant de la taille de leur ensemble d'apprentissage.

**Apprentissage par transfert et apprentissage tout au long de la vie :** Pour adapter chaque approche neuronale à un nouveau domaine (i.e., à un nouveau corpus), nous avons effectué un *fine-tuning* ( $\mathcal{D}$  dans la figure 1). De façon parallèle, nous avons utilisé la méthode *Elastic Weight Consolidation* (EWC) (Kirkpatrick *et al.*, 2016) comme méthode d'apprentissage tout au long de la vie afin d'évaluer son apport contre l'oubli catastrophique. EWC contraint la fonction de coût des modèles avec un facteur additionnel d'oubli, que nous avons appliqué sur les 5 modèles considérés.

**Métriques :** Nous reportons dans ce résumé les résultats obtenus pour la métrique REM (*Remembering*), dérivée de la mesure BWT (*backward transfer measure*) proposée par Rodriguez et al. (Díaz-Rodríguez *et al.*, 2018). Cette mesure évalue l'effet intrinsèque que l'apprentissage d'un modèle  $M$  sur un nouveau corpus (flèches performance à droite sur la figure 1) a sur la performance obtenue sur un précédent corpus (flèches performance à gauche sur la figure). Une valeur de 1 sur cette mesure signifie que le modèle ne souffre pas d'oubli catastrophique.

**Implémentation:** Le code des expérimentations est disponible ici : <https://github.com/jeslev/OpenNIR-Lifelong>. Il utilise le framework OpenNIR.

## 2.2. Résultats principaux

Le tableau 1 résume nos résultats. Afin de mieux mesurer la capacité intrinsèque des modèles d'ordonnancement neuronal à se souvenir d'une connaissance préalablement acquise, nous avons considéré pour chaque configuration expérimentale 2

**Tableau 1.** Résultats - REM(REMN) - par modèle pour la mesure REM (REMN) selon deux stratégies : fine-tuning et apprentissage tout au long de la vie (EWC). D'autres métriques sont disponibles dans (Lovón-Melgarejo et al., 2021).

Modèle	Configuration	Fine-tuning	EWC
DRMM	$ms \rightarrow c19$	1.000(1.000)	1.000(1.000)
	$ms \rightarrow mb$	0.962(0.943)	0.971(0.974)
	$mb \rightarrow c19$	1.000(0.965)	1.000(0.662)
	$ms \rightarrow mb \rightarrow c19$	0.976(0.938)	0.979(1.000)
PACRR	$ms \rightarrow c19$	1.000(0.760)	1.000(0.756)
	$ms \rightarrow mb$	1.000(1.000)	1.000(1.000)
	$mb \rightarrow c19$	1.000(0.523)	1.000(0.940)
	$ms \rightarrow mb \rightarrow c19$	1.000(0.759)	1.000(0.874)
KNRM	$ms \rightarrow c19$	1.000(1.000)	1.000(1.000)
	$ms \rightarrow mb$	1.000(1.000)	1.000(1.000)
	$mb \rightarrow c19$	1.000(0.810)	1.000(0.902)
	$ms \rightarrow mb \rightarrow c19$	1.000(1.000)	1.000(0.963)
VBERT	$ms \rightarrow c19$	0.930(1.000)	1.000(1.000)
	$ms \rightarrow mb$	1.000(0.883)	1.000(1.000)
	$mb \rightarrow c19$	0.913(1.000)	1.000(1.000)
	$ms \rightarrow mb \rightarrow c19$	0.989(0.922)	1.000(1.000)
CEDR	$ms \rightarrow c19$	0.826(1.000)	1.000(1.000)
	$ms \rightarrow mb$	0.510(0.920)	1.000(1.000)
	$mb \rightarrow c19$	0.940(1.000)	1.000(1.000)
	$ms \rightarrow mb \rightarrow c19$	0.771(0.946)	0.891(1.000)

scénarios : le premier combine le score du modèle neuronal avec un score BM25, et le second ne prend en compte que le score du modèle neuronal. Les résultats associés à ces variations correspondent respectivement aux métriques REM et REMN du tableau 1. Plus de détails et une discussion sont disponibles dans (Lovón-Melgarejo et al., 2021). Nous pouvons mettre en avant les conclusions suivantes :

- Nous avons montré l'existence de l'oubli catastrophique dans les modèles neuronaux pour l'ordonnancement des documents (colonne *Fine-tuning* du tableau 1 - **RQ1**). Plus le modèle d'ordonnancement est efficace de façon intrinsèque, plus les risques d'oubli catastrophique sont grands. Des modèles *transformers*, tels que CEDR et VBERT, permettent d'obtenir un bon compromis entre efficacité et oubli, comme démontré précédemment en TAL (Mosbach et al., 2020).

- Nous avons identifié des caractéristiques des corpus comme prédictors d'un oubli catastrophique (**RQ2**). A titre d'exemple, la difficulté d'un corpus impacte positivement la mémoire des modèles (c'est à dire diminue l'oubli catastrophique).

- Nous avons montré l'intérêt d'intégrer une méthode d'apprentissage tout au long de la vie pour atténuer l'oubli catastrophique (colonne *EWC* du tableau 1 **RQ3**).

## Remerciements

Les projets ANR COST (ANR-18-CE23-0016) and ANR JCJC SESAMS (ANR-18-CE23-0001) ont soutenu ce travail.

### 3. Bibliographie

- Asghar N., Mou L., Selby K. A., Pantasdo K. D., Poupart P., Jiang X., “Progressive Memory Banks for Incremental Domain Adaptation”, *ICLR*, vol. abs/1811.00239, 2020.
- Bajaj P., Campos D., Craswell N., Deng L., Gao J., Liu X., Majumder R., McNamara A., Mitra B., Nguyen T. *et al.*, “Ms marco: A human generated machine reading comprehension dataset”, *arXiv preprint arXiv:1611.09268*, 2016.
- Chen Z., Liu B., *Lifelong Machine Learning, Second Edition*, Synthesis Lectures on Artificial Intelligence and Machine Learning, n.d.
- de Masson d’Autume C., Ruder S., Kong L., Yogatama D., “Episodic Memory in Lifelong Language Learning”, *CoRR*, 2019.
- Díaz-Rodríguez N., Lomonaco V., Filliat D., Maltoni D., “Don’t forget, there is more than forgetting: new metrics for Continual Learning”, *ArXiv e-prints*, 2018.
- Guo J., Fan Y., Ai Q., Croft W. B., “A Deep Relevance Matching Model for Ad-Hoc Retrieval”, 2016.
- Hui K., Yates A., Berberich K., de Melo G., “Pacrr: A position-aware neural ir model for relevance matching”, *arXiv preprint arXiv:1704.03940*, 2017.
- Kirkpatrick J., Pascanu R., Rabinowitz N. C., Veness J., Desjardins G., Rusu A. A., Milan K., Quan J., Ramalho T., Grabska-Barwinska A., Hassabis D., Clopath C., Kumaran D., Hadsell R., “Overcoming catastrophic forgetting in neural networks”, *CoRR*, 2016.
- Li Z., Hoiem D., “Learning without Forgetting”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.
- Lovón-Melgarejo J., Soulier L., Pinel-Sauvagnat K., Tamine L., “Studying Catastrophic Forgetting in Neural Ranking Models”, *Proceedings of ECIR 2021, to appear*, 2021.
- MacAvaney S., Yates A., Cohan A., Goharian N., “CEDR: Contextualized embeddings for document ranking”, *ACM SIGIR*, p. 1101-1104, 2019.
- Mosbach M., Andriushchenko M., Klakow D., “On the Stability of Fine-tuning BERT: Misconceptions, Explanations, and Strong Baselines”, 2020.
- Pan S. J., Yang Q., “A Survey on Transfer Learning”, 2010.
- Parisi G. I., Kemker R., Part J. L., Kanan C., Wermter S., “Continual lifelong learning with neural networks: A review”, *Neural Networks*, vol. 113, p. 54-71, 2019.
- Thompson B., Gwinnup J., Khayrallah H., Duh K., Koehn P., “Overcoming Catastrophic Forgetting During Domain Adaptation of Neural Machine Translation”, *NAACL*, p. 2062-2068, 2019.
- Wang L. L., Lo K., Chandrasekhar Y., Reas R., Yang J., Eide D., Funk K., Kinney R., Liu Z., Merrill W. *et al.*, “CORD-19: The Covid-19 Open Research Dataset”, *ArXiv*, 2020.
- Wiese G., Weissenborn D., Neves M., “Neural Domain Adaptation for Biomedical Question Answering”, *CoNLL 2017*, August, 2017.
- Xiong C., Dai Z., Callan J., Liu Z., Power R., “End-to-end neural ad-hoc ranking with kernel pooling”, *ACM SIGIR*, p. 55-64, 2017.
- Yang W., Lu K., Yang P., Lin J., “Critically Examining the "Neural Hype" Weak Baselines and the Additivity of Effectiveness Gains from Neural Ranking Models”, *ACM SIGIR*, p. 1129-1132, 2019.