Détecter des comportements associés aux troubles alimentaires par l'analyse automatique des publications textuelles en ligne

Yves Ferstler Catherine Lavoie Marie-Jean Meurs
Université du Québec à Montréal, Montréal, QC, Canada
ferstler.yves, lavoie.catherine.15}@courrier.uqam.ca,
meurs.marie-jean@uqam.ca

RÉSUMÉ

Cet article présente une méthode pour détecter des aspects du comportement liés aux troubles alimentaires à partir de publications textuelles échangées sur les réseaux sociaux. Nos travaux comparent différentes représentations d'historiques de publications permettant d'entraîner un modèle neuronal pour la prédiction. Les approches étudiées sont : (1) la représentation de sujet par fréquence, en calculant le nombre de sujets apparus dans un historique, (2) une représentation par plongement, en calculant la moyenne des représentations de sujets présents dans l'historique de publications, (3) une représentation par documents représentatifs, qui cherche à représenter un sujet par un document sémantiquement proche. Un filtrage de sujets est également étudié, pour sélectionner les sujets reliés aux troubles alimentaires. Les résultats montrent que l'utilisation de filtrage permet d'améliorer les performances des systèmes de détection. La méthode basée sur un document représentatif obtient les meilleurs résultats, parmi les autres représentations évaluées mais également parmi d'autres méthodes appliquées à la même tâche lors de la campagne d'évaluation eRisk 2024.

ABSTRACT

Detecting signs of eating disorders through automatic analysis of online text conversations

This paper leverages a topic modeling algorithm to predict aspects of eating disorders from a dataset of social network publications. Specifically, it investigates three methods for representing users' histories : (1) a frequency-based approach that calculates topic frequencies across a user's history, providing a quick but potentially insightful representation. (2) an embedding-based approach that uses a cluster topic representation where two clustering algorithms dense and centroid-based are tested to enrich semantic information. (3) a document-based approach that refines cluster embeddings by identifying the single most representative document, thereby mitigating the limitations observed in the embedding-based approach. A filter approach is also evaluated, focused on documents that share similarities with a standard questionnaire for self-evaluation of eating disorders. The results indicate that all representative approaches with filtering improved performance. Among the three representation methods, the representative document achieves the highest scores. Compared to other approaches applied on the same task, this method shows better results in 4 over 8 metrics compared to all other approaches proposed by the eRisk 2024 evaluation campagn participants.

MOTS-CLÉS : Modèle de sujet, Troubles alimentaires, Représentation d'historique conversationnel.

KEYWORDS: Topic model, Eating disorders, Representation of user history of publications.

1 Introduction

Les troubles alimentaires (TA) sont complexes et ont de nombreux impacts sur la vie d'une personne et son bien-être. Caractérisés par une alimentation anormale culturellement, des craintes nutritionnelles, ou des préoccupations persistantes liées à l'image corporelle (Crocq *et al.*, 2015), les TA sont variés. De l'anorexie mentale à la boulimie ou encore les accès hyperphagiques, chacun de ces troubles montre des comportements associés variables et des symptômes différents. Plusieurs facteurs génétiques, psychologiques, sociaux ou culturels, peuvent favoriser le développement de troubles alimentaires chez une personne (Crocq *et al.*, 2015). Les symptômes pouvant être difficiles à détecter et confondus avec ceux d'autres troubles, l'identification de TA chez une personne atteinte peut se révéler complexe. C'est pour cette raison que des questionnaires d'auto-évaluation (Aardoom *et al.*, 2012a) sont utilisés dans le processus diagnostique.
Il est aujourd'hui courant de partager ses sentiments ou ses expériences en ligne : communiquer avec d'autres personnes, partager ses préoccupations et chercher de l'aide sur internet est devenu une habitude pour de nombreuses personnes (Sowles *et al.*, 2018). Des corpora de messages et de discussions entre internautes sont disponibles pour la communauté de recherche et peuvent être utilisés pour des projets liant le domaine du traitement automatique du langage naturel (TALN) et celui de la santé mentale. L'analyse automatique des conversations textuelles peut en effet aider à identifier des comportements associés aux troubles mentaux, et notamment de TA. Pour entraîner un réseau neuronal sur des tâches textuelles, une méthode de représentation des données doit être choisie au préalable.
Dans ce travail, l'approche utilisée pour représenter le texte brut en message interprétable est l'utilisation d'un modèle de sujet, qui permet d'identifier les différents sujets abordés par une personne utilisatrice dans un historique conversationnel.
Dans un premier temps nous nous intéressons à la représentation de l'historique des publications par le modèle de sujet, en utilisant les méthodes de fréquences, de plongement et de document représentatif sémantiquement proche. Dans un second temps, nous examinons comment utiliser ces représentations et si une sélection de sujets permet d'améliorer les performances des systèmes de détection proposés.

2 État de l'art

Selon l'organisation mondiale de la santé, en 2022, une personne sur 8 dans le monde était affectée par un ou plusieurs troubles mentaux et 14 millions de personnes dont 3 millions d'enfants et adolescents avaient déjà souffert de TA [1]. L'aide que peut apporter le TALN aux recherches dans le domaine de la santé mentale et plus spécifiquement des TA est donc essentielle pour tenter d'aider les personnes qui souffrent. Parmi les travaux récents en TALN, Chiong *et al.* (2021) proposent une méthode de détection des signes de dépression sur les réseaux sociaux en utilisant des techniques de correction de mots, d'étiquetage Parts-Of-Speech (POS) et de lemmatisation, afin d'entraîner un modèle de classification. D'autres techniques utilisent des transformeurs puis des modèles d'apprentissage pour détecter des signes de cyberharcèlement et de comportements associés aux TA (Sihab-Us-Sakib *et al.*, 2024; Karamat *et al.*, 2024).

Rujas *et al.* (2023) présentent une méthode qui combine le prétraitement de données avec la modélisation de sujet en utilisant BERTopic (Grootendorst, 2022), et qui obtient des résultats prometteurs pour

1. OMS, 2022 https://www.who.int/news-room/fact-sheets/detail/mental-disorders

la détection précoce des TA à partir de messages textuels. Benitez-Andrades *et al.* (2021) proposent une comparaison de six modèles de langue préentrainés – BERT (Devlin *et al.*, 2019), RoBERTa (Liu *et al.*, 2019), DistilBERT (Sanh *et al.*, 2020), CamemBERT (Martin *et al.*, 2020), ALBERT (Lan *et al.*, 2020), FlauBERT (Le *et al.*, 2020) – pour détecter les comportements liés aux TA sur les réseaux sociaux. Ces modèles obtiennent des résultats intéressants, particulièrement avec l'utilisation de RoBERta. Tout en limitant la surconsommation de ressources computationnelles, des modèles plus frugaux comme ceux de Wang *et al.* (2020) obtiennent de bons résultats autant en termes d'efficacité que de performance.
Les travaux présentés précédemment indiquent que l'emploi de modèles de sujet pour représenter des historiques conversationnels est un choix pertinent. Parmi ces modèles, BERTopic permet d'obtenir des résultats à l'état de l'art dans le domaine de la détection du risque en santé mentale. Dans le cadre de la campagne d'évaluation eRisk 2024 [2], les équipes participant à la tâche 3 ont travaillé sur l'évaluation de signes de troubles alimentaires, et parmi les résultats (Parapar *et al.*, 2024), l'approche utilisant BERTopic (Maupomé *et al.*, 2024) a obtenu sur plusieurs métriques des résultats proches des meilleurs (Prasanna *et al.*, 2024). Cette équipe a obtenu les meilleurs résultats à partir de 3 méthodes différentes, elles combinent des techniques de Word2Vec, de rétrotraduction, de réduction de dimensions (PCA) et de classifieurs SVM.

3 Méthodologie

Notre objectif est de représenter les historiques de publications en données interprétables pour l'entraînement d'un réseau neuronal, tout en préservant leur richesse informationnelle. Nos travaux ont donc porté sur le choix de l'ensemble de données et la représentation des publications et des historiques pour entraîner le modèle prédictif.

3.1 Ensemble de données

Le choix d'un ensemble de données dans un contexte de TA nécessite de porter une attention particulière à l'équilibre des données, à la présence de faux positifs et de faux négatifs. L'ensemble de données choisi dans les travaux présentés est celui de la tâche 3 de la conférence eRisk 2024, qui propose un ensemble de données issues de réseaux sociaux. La tâche de 2024 est la même que celles de 2022 et 2023. L'ensemble est constitué des données de ces deux années pour l'entraînement du modèle, et des données de test de 2024 pour l'évaluation. Ces ensembles contiennent l'historique conversationnel d'une personne utilisatrice, associée à ses réponses au questionnaire EDE-Q (*The Eating Disorder Examination Questionnaire*) (Fairburn & Beglin, 2008), un questionnaire standard d'auto-évaluation sur les comportements généralement associés aux TA.

Le questionnaire EDE-Q comporte 28 questions visant à évaluer les comportements associés à certains troubles alimentaires durant une période de temps précise. Les réponses varient sur une échelle de 0 à 6, où **0** indique que le ou les comportements associés aux TA évalués dans la question ne se sont jamais manifestés, et **6** indique qu'ils se sont produits quotidiennement. Ainsi, plus le score est élevé, plus forte est la prévalence du ou des comportements.

Quatre sous-échelles portent sur différents aspects des TA : la restriction alimentaire, les préoccupa-

2. eRisk 2024 https://erisk.irlab.org/2024/index.html

tions alimentaires, la silhouette, le poids. Le score global est obtenu en combinant les résultats des quatre sous-échelles et en faisant la moyenne des réponses attribuées à chaque question. Le tableau 1 fournit des exemples de questions présentes dans le EDE-Q et leur traduction libre en français.

#	Question	Sous-échelle
1.	*Have you been deliberately **trying** to limit the amount of food you eat to influence your shape or weight (whether or not you have succeeded) ?* – Avez-vous délibérément essayé de limiter la quantité de nourriture que vous mangez pour influencer votre forme ou votre poids (que vous ayez réussi ou non) ?	Restriction alimentaire
7.	*Has thinking about **food, eating or calories** made it very difcult to concentrate on things you are interested in (for example, working, following a conversation, or reading) ?* – Penser à la nourriture, à l'alimentation ou aux calories vous a-t-il empêché de vous concentrer sur des choses qui vous intéressent (par exemple, travailler, suivre une conversation ou lire) ?	Préoccupations alimentaires
6.	*Have you had a definite desire to have a **totally flat** stomach ?* – Avez-vous déjà eu le désir absolu d'avoir un ventre totalement plat ?	Silhouette
12.	*Have you had a strong desire to lose weight ?* – Avez-vous eu un fort désir de perdre du poids ?	Poids

TABLE 1 – Exemples de questions du EDE-Q avec leurs sous-échelles

Le score global et les scores des sous-échelles sont utilisés pour interpréter les réponses du questionnaire (Mond *et al.*, 2004b; Institute for Eating Disorders, 2024 12 31; Aardoom *et al.*, 2012b; Mond *et al.*, 2004a). Une interprétation courante du score global est qu'un score plus élevé reflète des comportements associés aux TA plus marqués (Aardoom *et al.*, 2012b), un score proche de 3 est également utilisé comme seuil pour différencier les cas positifs des cas négatifs (Institute for Eating Disorders, 2024 12 31). Parmi les 28 questions, seuls 22 sont utilisées pour la cotation des sous-échelles et du score global (les questions 13 à 18 sont exclues car les réponses attendues sont en texte libre).

Le corpus d'entraînement contient 74 historiques de conversations textuelles, chacun étant associé à une personne utilisatrice. Parmi les 74 personnes utilisatrices, la moitié montre des comportements associés aux TA, le corpus est donc équilibré. Une description détaillée est fournie dans le tableau 2. Les Figures 1 et 2 montrent en bleu le groupe présentant des comportements associés aux TA et en orange le groupe sans ces comportements, en utilisant le seuil de 3 du score global. Les tendances des scores entre les deux groupes sont similaires, que ce soit pour les sous-échelles ou pour les questions individuelles. Les personnes utilisatrices sans comportements associés aux TA partagent certaines préoccupations des personnes présentant ces comportements, mais avec une fréquence moindre. Une différence majeure entre les groupes réside dans les scores EDE-Q, qui sont proportionnellement plus élevés pour les personnes utilisatrices présentant des comportements associés aux TA.

Les statistiques du tableau 2 montrent un nombre de publications relativement élevé (32 806) pour un nombre de personnes utilisatrices faible (74). Avec une médiane de 196 publications et un troisième quartile de 912 publications par personne utilisatrice, une majorité de l'information de l'ensemble de

Statistiques		Mots les plus fréquents	
Métrique	**Valeur**	**Mot**	**Nb. occurences**
Nombre de personnes utilisatrices	74	***like***	7 214
Nombre de publications	32 806	***people***	4 319
Nombre de mots total	1 185 532	*get*	3 909
Taille du vocabulaire	104 757	***think***	3 470
Médiane de mots par publication	17,0	*would*	3 445
1er Quartile de mots par publication	6,0	*one*	3 142
3ème Quartile de mots par publication	43,0	*know*	3 098
Moyenne de publications par personne	443,32	*really*	2 938
Médiane de publications par personne	196,0	*time*	2 844
1er Quartile de publications par personne	86,5	***feel***	2 584
3ème Quartile de publications par personne	912,5	*much*	2 246
Moyenne de publications mensuelles par personne	144,15	*also*	2 219
Médiane de publications mensuelles par personnes	59,5	***want***	2 152

TABLE 2 – Description de l'ensemble de données d'entrainement de la tâche 3 de eRisk 2024

données sera contenu dans une minorité d'historiques conversationnels. Pour permettre au modèle neuronal d'observer une plus grande diversité d'historiques, nous avons segmenté les historiques originaux en portions d'historiques de 350 publications (au maximum). Les réponses originales aux EDE-Q sont associées à la portion d'historique contenant les publications les plus récentes. Pour éviter le sur-apprentissage, et sur recommandation de personnes cliniciennes, les autres portions d'historiques sont associées aux réponses EDE-Q originales, en ajoutant du bruit. Ainsi, pour 10% des réponses, la valeur est modifiée de $\pm$ 1 entre 1 et 5, de + 1 pour la valeur 0 et de -1 pour la valeur 6. Le corpus d'entraînement ainsi segmenté comporte 138 historiques conversationnels, majorés à 350 publication par historiques, avec une médiane de publications par personne passant de 196,0 à 335,0 et d'un 3ème quartile de publications par personne passant de 912,5 à 349,0.

3.2 Représentation

Représenter les caractéristiques des historiques sous une forme riche est essentiel pour obtenir du modèle neuronal la meilleure prédiction possible. Notre choix de représentation d'un historique comporte deux étapes : (1) la transformation des publications en représentation individuelles, puis (2) l'agrégation des représentations individuelles pour représenter complètement l'historique.

3.2.1 Représentation des publications

Nous avons choisi d'explorer une représentation par modélisation de sujet en utilisant BERTopic (Grootendorst, 2022), qui génère, lors de son l'entraînement, un sujet pour chaque publication. Les publications sémantiquement proches seront donc associées au même sujet et les sujets générés pendant l'entraînement seront utilisés pour étiqueter les nouvelles publications pendant la phase de test. BERTopic fournit également trois publications représentatives pour chaque sujet.

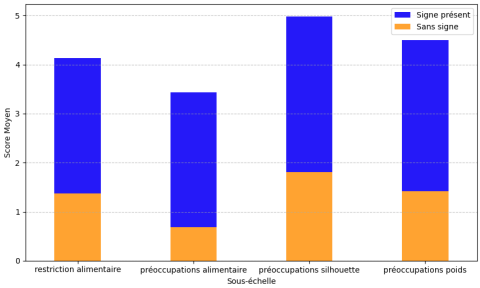

FIGURE 1 – Score moyen des sous-échelles entre les personnes utilisatrices ayant des comportements associés aux TA et celles sans.

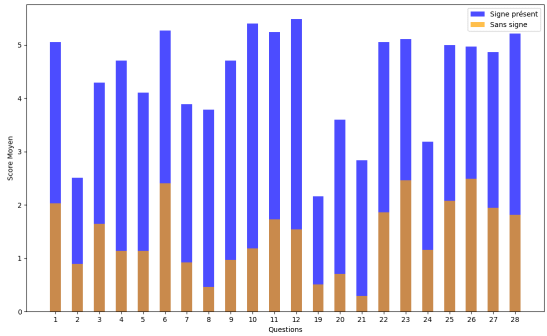

FIGURE 2 – Score moyen des questions entre les personnes ayant des comportements associés aux TA et celles sans.

Nous avons choisi de représenter chaque sujet de trois manières différentes : soit par une valeur numérique, soit par son plongement dans BERTopic, soit par sa publication représentative la plus proche de la nouvelle publication à étiqueter. Dans ce dernier cas, le choix d'une publication représentative parmi les trois fournies par BERTopic se fait dynamiquement et la distance utilisée est la distance euclidienne.

Pour générer les sujets, BERTopic utilise HDBSCAN (Malzer & Baum, 2020), un algorithme de clustering hiérarchique basé sur des zones de forte densité des points représentant les documents, puis calcule le centroïde des clusters obtenus. Tel qu'illustré à la figure 3, cette approche peut être problématique. En effet, bien que deux clusters distincts puissent être correctement identifiés, il est possible que leurs sujets soient confondus.

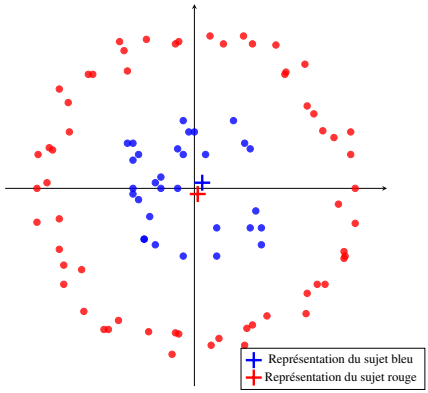

FIGURE 3 – Exemple de scénario problématique : deux clusters (un rouge, un bleu) et leurs centroïdes sous forme de plus (+) situées à des coordonnées voisines.

Pour éviter ce problème, nous avons aussi utilisé l'agorithme K-means car il génère des clusters et donc des centroïdes topologiquement éloignés. Le nombre de clusters imposé à K-means est le même que le celui déterminé par HDBSCAN, soit 413.

3.2.2 Représentation des historiques

Trois représentations d'historique sont évaluées : une basée sur la fréquence des sujets présents et deux basées sur les plongements – par publications et par documents représentatifs.
L'approche basée sur la fréquence représente l'historique d'une personne utilisatrice en calculant la fréquence d'apparition de chaque sujet.Cette représentation est efficace en termes de calcul et facilement interprétable puisqu'elle associe une liste de sujets à un historique. Cependant, cette simplicité limite la richesse de l'information encodée, ce qui peut affecter la performance prédictive du modèle neuronal. De plus, lorsque l'ensemble des sujets est large ou que le nombre de publications dans l'historique est faible, la représentation peut contenir une proportion significative de valeurs nulles, réduisant la performance du modèle neuronal.

L'approche basée sur la moyenne permet de représenter l'historique en calculant la moyenne des plongements associés aux publications ou la moyenne des plongements associés aux documents représentatifs. La taille de cette représentation est identique à celle des plongements utilisés. Cette méthode favorise une plus grande richesse de représentation.

3.2.3 Filtrage

Quelle que soit la technique de représentation, toutes les informations extraites ne sont pas forcément pertinentes dans le contexte de la détection des comportements associés aux TA. Ainsi, les sujets générés reflètent la globalité des conversations et l'utilisation de tous les sujets peut être donc inadaptée à la détection des comportements associés aux TA. Par conséquent, un filtrage des publications est appliqué afin d'identifier les plus pertinentes pour la représentation de l'historique. Pour cela, une sélection de sujets est faite regroupant les sujets attribués à chaque question du questionnaire EDE-Q par BERTopic. Afin d'élargir cette sélection, des sujets sémantiquement proches de ceux associés aux questions sont également ajoutés à la liste. Un ratio est utilisé pour définir la proportion de sujets à filtrer. Comme ce ratio est appliqué indépendamment à chaque question, il ne reflète pas directement le nombre total de sujets filtrés.

3.2.4 Modèle de prédiction de l'EDE-Q

Le modèle de prédiction pour le questionnaire EDE-Q est inspirée de (Maupomé *et al.*, 2024). Ce modèle utilise trois réseaux neuronaux à action direct (*feed-forward networks*). Le premier réseau, entraîné à partir des représentations d'historiques conversationnels, prédit le score global obtenu au questionnaire. Le second réseau, entraîné à partir du modèle du score global et des représentations d'historiques conversationnels, prédit les scores des sous-échelles. Enfin, le troisième réseau, entraîné à partir des modèles du score global, des sous-échelles et des représentations d'historiques conversationnels, prédit les réponses aux questions.

4 Expériences et résultats

4.1 Métriques

Les métriques utilisées pour mesurer la performance des approches proposées sont celles de la campagne d'évaluation eRisk 2024, ce qui permet de comparer nos résultats à ceux des équipes participantes. Ces métriques sont les suivantes :

- **Mean Zero-One Error (MZOE)** mesure la proportion de réponses des personnes utilisatrices pour lesquelles la réponse prédite est incorrecte, peu importe le degré de l'erreur.
- **Mean Absolute Error (MAE)** est la moyenne sur tous les historiques des écarts moyens entre la réponse prédite et la réponse réelle des personnes utilisatrices.
- **Macro-averaged MAE (MAE_macro)** est la moyenne des MAE sur tous les historiques des écarts moyens pour chaque catégorie de réponse (définie par un score allant de 0 à 6).
- **Restriction (RS, questions 1 à 5), Préoccupation alimentaire (ECS, questions 7, 9, 19, 21, 20), Préoccupation de la silhouette (SCS questions 6, 8, 10, 11, 23, 26, 27, 28) et Préoccupation du poids (WCS, questions 8, 12, 22, 24, 25)** se concentrent sur des sous-échelles spécifiques des troubles alimentaires en calculant la racine de l'écart quadratique moyen entre la moyenne des scores prédits par le modèle et ceux des vraies réponses sur les questions associées.
- **Score global (GED)** est la même métrique appliquée sur les quatre sous-échelles précédentes afin de fournir une évaluation globale des performances prédictives sur l'ensemble du questionnaire.

4.2 Contrainte temporelle imposée par le questionnaire EDE-Q

Le EDE-Q 6.0 déclare que les questions doivent être répondues selon les évènements des 28 derniers jours (Fairburn & Beglin, 2008). Cependant, la proportion de données datant des 28 derniers jours est minimale comparée au corpus complet. L'utilisation de filtrage de sujets réduisant la taille de l'ensemble de données, une restriction supplémentaire du corpus aux publications des 28 jours les plus récents limite la capacité d'entraîner le modèle neuronal. Nous avons donc examiné si les publications plus anciennes pourraient aider à identifier des comportements associés aux TA en comparant les résultats obtenus sur le sous-corpus contenant uniquement les publications des 28 derniers jours et sur un corpus sans restriction temporelle. Afin d'assurer une répartition équitable du nombre de publications entre les deux corpus, le nombre de publications du second est réduit à celui du premier en supprimant de manière aléatoire certaines publications de l'historique complet. Les résultats obtenus montrent que l'utilisation du corpus sans restriction de date et contenant la même quantité de données permet d'obtenir de meilleures prédictions que le corpus restreint à 28 jours. Bien que le questionnaire EDE-Q exige que les personnes participantes répondent aux questions en fonction de leurs comportements des derniers 28 jours, il est donc pertinent de considérer les historiques conversationnels des personnes sur une période non contrainte.

4.3 Expériences

Les expériences comparent les trois types de représentation : par fréquence, par plongement et par document représentatif le plus proche. Les configurations étudiées font varier les paramètres de filtrage de sujets. HDBSCAN et K-means sont comparés pour la représentation par plongement.

	Sans filtrage	Filtrage moyen	Filtrage élevé	Filtrage fort
K-means	441	231	144	83
HDBSCAN	412	224	148	85

TABLE 3 – Proportion de sujets générés par BERTopic en fonction du filtrage est présentée sur le tableau.

run eRisk	MAE	MZOE	MAE_macro	GED	RS	ECS	SCS	WCS
baseline tous 0	3,790	0,813	4,254	4,472	3,869	4,479	4,363	3,361
baseline tous 6	1,937	**0,551**	3,018	3,076	3,352	2,868	3,029	2,472
baseline moyenne	1,965	0,594	3,137	2,875	3,361	2,102	2,229	2,306
RELAI_0 (Maupomé *et al.*, 2024)	2,331	0,914	2,243	2,394	2,222	2,324	2,340	**1,812**
SCaLAR-NITK_0 (Prasanna *et al.*, 2024)	1,912	0,591	**1,643**	2,495	2,713	1,568	**1,536**	2,098
SCaLAR-NITK_1	1,980	0,664	1,972	2,570	2,562	**1,553**	1,960	2,066
SCaLAR-NITK_2	1,879	**0,568**	1,942	2,158	2,477	2,222	2,245	2,364
SCaLAR-NITK_3	1,932	0,586	1,868	**2,117**	2,430	2,046	2,242	2,407
SCaLAR-NITK_4	**1,874**	0,672	1,820	2,292	**2,140**	1,557	1,880	2,061
Représentations évaluées								
Fréquence	**1,775**	0,747	1,692	**1,967**	**2,024**	2,003	1,896	1,942
Plongement	1,840	0,767	1,727	2,126	2,103	2,217	2,067	2,112
Document représentatif	**1,694**	0,739	**1,620**	**2,061**	**2,013**	2,184	1,989	2,052

TABLE 4 – Baseline, meilleurs résultats obtenus à la tâche 3 de eRisk 2024 et résultats obtenus par nos trois représentations. Un score **bas** montre de meilleurs résultats pour l'ensemble des métriques.

Trois filtrages sont appliqués sur les sujets pour conserver : les 10% les plus pertinents (filtrage fort), les 25% les plus pertinents (filtrage élevé) et les 50% les plus pertinents (filtrage moyen). Une configuration sans filtrage est également testée. La proportion de sujets générés par BERTopic en fonction du filtrage est présentée dans le tableau 3.

BERTopic utilise le plongement all-MiniLM-L6-v2 [3] pour chacune des représentations, et HDBSCAN est utilisé comme algorithme de clustering pour les représentations de fréquence et de document représentatifs. Les modèles ont été entraînés et validés à partir de l'ensemble de données d'entraînement (124 historiques) et de validation (14 historiques) de la campagne eRisk 2022-2023. L'ensemble de test (18 historiques) est celui de eRisk 2024.

4.4 Résultats

Les configurations ayant obtenu les meilleurs résultats sont :

— HDBSCAN et filtrage fort (∼10%) pour la représentation par fréquence.
— K-means et filtrage moyen (∼ 50%) pour la représentation par plongement.
— HDBSCAN et filtrage élevé (∼ 25%) pour la représentation par document représentatif.

Les résultats de l'évaluation des trois représentations sont présentés dans le bas du tableau 4. La

3. https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2

représentation par document représentatif obtient les meilleurs résultats pour 4 des 8 métriques (MZOE, MAE, MAE_macro et RS). La représentation par fréquence montre les meilleurs résultats pour chacune des sous-échelles et pour le score global.

Comparée avec les meilleurs scores obtenus à la tâche 3 d'eRisk 2024 présentés en haut du tableau 4, la représentation par document représentatif obtient des meilleurs scores sur 4 métriques (MAE, MAE_macro, GED et RS). Les résultats des représentations par fréquence et par plongement restent intéressants en comparaison des résultats obtenus à eRisk 2024.

Les résultats confortent dans le choix d'utiliser un modèle de sujet pour représenter les publications des historiques. L'approche par document représentatif confirme qu'une représentation plus riche des données conduit à de meilleurs résultats.
La représentation basée sur la fréquence produit également des résultats intéressants. Associée à un filtrage fort, elle semble avoir empêché l'extraction d'attributs non pertinents. Enfin, bien que la représentation par plongement présente des performances légèrement inférieures aux deux autres méthodes, ses résultats restent proches de l'état de l'art.

5 Conclusion

Ces travaux comparent trois méthodes de représentation d'historiques conversationnels pour prédire des comportement associés aux TA : représentation par fréquence, par plongement et par document représentatif. Un filtrage de sujets, effectué à partir des questions du EDE-Q, est également évalué dans la représentation d'historique. Cette représentation, une fois obtenue, permet d'entraîner un modèle neuronal pour la prédiction. Les résultats montrent que le filtrage permet d'améliorer la prédiction pour chacune des représentations. L'approche par document représentatif présente les meilleurs résultats, supérieurs à l'état de l'art de la tâche sur quatre métriques .

Nos travaux en cours se penchent sur le choix de l'algorithme de cluster permettant la création de sujets, les méthodes de filtrage possibles des listes de sujets et l'évaluation de techniques de réduction de dimension bien adaptées à la modélisation de sujet. Les travaux sont menés en collaboration avec des personnes expertes en santé mentale, ce qui permet d'explorer également l'ajout potentiel de ressources cliniques externes pour enrichir les outils développés.

Pour assurer la reproductibilité des expériences, le code source est disponible sous licence libre ici : https://gitlab.labikb.ca/ikb-lab/articles/taln-2025.

Remerciements

Ces travaux ont été réalisés grâce aux ressources de calcul mises à notre disposition par Calcul Québec et l'Alliance de recherche numérique du Canada, et grâce au soutien financier du Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG) [MJ Meurs, CRSNG à la découverte #06487-2017] et de la Chaire de recherche du Québec sur la découvrabilité des contenus scientifiques en français [MJ Meurs, DOI 10.69777/358425]. Nous souhaitons également remercier Yassine Chahdi pour sa participation enthousiaste aux réflexions et aux analyses post-évaluation.

Références

AARDOOM J. J., DINGEMANS A. E., SLOF OP'T LANDT M. C. & VAN FURTH E. F. (2012a). Norms and discriminative validity of the Eating Disorder Examination Questionnaire (EDE-Q). *Eating Behaviors*, **13**(4), 305–309. DOI : 10.1016/j.eatbeh.2012.09.002.

AARDOOM J. J., DINGEMANS A. E., SLOF OP'T LANDT M. C. & VAN FURTH E. F. (2012b). Norms and discriminative validity of the Eating Disorder Examination Questionnaire (EDE-Q). *Eating Behaviors*, **13**(4), 305–309. DOI : https ://doi.org/10.1016/j.eatbeh.2012.09.002.

BENAMARA F., HATOUT N., MULLER P. & OZDOWSKA S., Éds. (2007). *Actes de TALN 2007 (Traitement automatique des langues naturelles)*, Toulouse. ATALA, IRIT.

BENITEZ-ANDRADES J. A., ALIJA-PEREZ J. M., GARCIA-RODRIGUEZ I., BENAVIDES C., ALAIZ-MORETON H., VARGAS R. P. & GARCIA-ORDAS M. T. (2021). BERT Model-Based Approach For Detecting Categories of Tweets in the Field of Eating Disorders (ED). In *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)*, p. 586–590, Aveiro, Portugal : IEEE. DOI : 10.1109/CBMS52027.2021.00105.

CHIONG R., BUDHI G. S., DHAKAL S. & CHIONG F. (2021). A textual-based featuring approach for depression detection using machine learning classifiers and social media texts. *Computers in Biology and Medicine*, **135**, 104499. DOI : https ://doi.org/10.1016/j.compbiomed.2021.104499.

CROCQ M.-A., GUELFI J. D., ASSOCIATION A. P. & FORCE A. P. A. D.-. T. (2015). *DSM-5 : manuel diagnostique et statistique des troubles mentaux (5e édition)*. Elsevier Masson.

DEVLIN J., CHANG M.-W., LEE K. & TOUTANOVA K. (2019). BERT : Pre-training of Deep Bidirectional Transformers for Language Understanding.

DIAS G., Éd. (2015). *Actes de TALN 2015 (Traitement automatique des langues naturelles)*, Caen. ATALA, HULTECH.

FAIRBURN C. & BEGLIN S. (2008). Eating Disorder Examination Questionnaire (EDE-Q 6.0). *Cognitive behavior therapy and eating disorders*, p. 309–313. DOI : 10.1037/t03974-000.

GROOTENDORST M. (2022). BERTopic : Neural topic modeling with a class-based TF-IDF procedure. DOI : https ://doi.org/10.48550/arXiv.2203.05794.

INSTITUTE FOR EATING DISORDERS (2024-12-31). Eating Disorder Examination Questionnaire EDE-Q. https://insideoutinstitute.org.au/resource-library/eating-disorder-examination-questionnaire-ede-q. Accessed : 2024-12-31.

KARAMAT A., IMRAN M., YASEEN M. U., BUKHSH R., ASLAM S. & ASHRAF N. (2024). A Hybrid Transformer Architecture for Multiclass Mental Illness Prediction using Social Media Text. *IEEE Access*, p. 1–1. DOI : 10.1109/ACCESS.2024.3519308.

LAIGNELET M. & RIOULT F. (2009). Repérer automatiquement les segments obsolescents à l'aide d'indices sémantiques et discursifs. In A. NAZARENKO & T. POIBEAU, Éds., *Actes de TALN 2009 (Traitement automatique des langues naturelles)*, Senlis : ATALA LIPN.

LAN Z., CHEN M., GOODMAN S., GIMPEL K., SHARMA P. & SORICUT R. (2020). ALBERT : A Lite BERT for Self-supervised Learning of Language Representations.

LANGLAIS P. & PATRY A. (2007). Enrichissement d'un lexique bilingue par analogie. In (Benamara *et al.*, 2007), p. 101–110.

LE H., VIAL L., FREJ J., SEGONNE V., COAVOUX M., LECOUTEUX B., ALLAUZEN A., CRABBÉ B., BESACIER L. & SCHWAB D. (2020). FlauBERT : Unsupervised Language Model Pre-training for French.

LIU Y., OTT M., GOYAL N., DU J., JOSHI M., CHEN D., LEVY O., LEWIS M., ZETTLEMOYER L. & STOYANOV V. (2019). RoBERTa : A Robustly Optimized BERT Pretraining Approach.

MALZER C. & BAUM M. (2020). A Hybrid Approach To Hierarchical Density-based Cluster Selection. In *2020 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*, p. 223–228 : IEEE. DOI : 10.1109/mfi49285.2020.9235263.

MARTIN L., MULLER B., ORTIZ SUÁREZ P. J., DUPONT Y., ROMARY L., DE LA CLERGERIE É., SEDDAH D. & SAGOT B. (2020). CamemBERT : a Tasty French Language Model. In D. JURAFSKY, J. CHAI, N. SCHLUTER & J. TETREAULT, Éds., *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, p. 7203–7219, Online : Association for Computational Linguistics. DOI : 10.18653/v1/2020.acl-main.645.

MAUPOMÉ D., FERSTLER Y., MOSSER S. & MEURS M.-J. (2024). Automatically finding evidence, predicting answers in mental health self-report questionnaires. In *eRisk 2024 Workshop at the 15th International Conference of the CLEF Association, CLEF 2024, Grenoble, France, September 9–12, 2024*, p. 841–850.

MOND J., HAY P., RODGERS B., OWEN C. & BEUMONT P. (2004a). Validity of the Eating Disorder Examination Questionnaire (EDE-Q) in screening for eating disorders in community samples. *Behaviour Research and Therapy*, **42**(5), 551–567. DOI : https ://doi.org/10.1016/S0005-7967(03)00161-X.

MOND J. M., HAY P. J., RODGERS B., OWEN C. & BEUMONT P. J. (2004b). Validity of the Eating Disorder Examination Questionnaire (EDE-Q) in screening for eating disorders in community samples. *Behaviour research and therapy*, **42**(5), 551–567.

PARAPAR J., MARTÍN-RODILLA P., LOSADA D. E. & CRESTANI F. (2024). Overview of eRisk 2024 : Early Risk Prediction on the Internet. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction : 15th International Conference of the CLEF Association, CLEF 2024, Grenoble, France, September 9–12, 2024, Proceedings, Part II*, p. 73–92, Berlin, Heidelberg : Springer-Verlag. DOI : 10.1007/978-3-031-71908-0_4.

PRASANNA S., GULATI A. S., KARMAKAR S., HIRANMAYI M. Y. & MADASAMY A. K. (2024). Measuring the severity of the signs of eating disorders using machine learning techniques. In *Experimental IR Meets Multilinguality, Multimodality, and Interaction : 15th International Conference of the CLEF Association, CLEF 2024, Grenoble, France, September 9–12, 2024, Proceedings, Part II*, p. 881–887.

RUJAS M., MERINO-BARBANCHO B., ARROYO P. & FICO G. (2023). Development of a Natural Language Processing-Based System for Characterizing Eating Disorders. In *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023)* : CEUR-WS.

SANH V., DEBUT L., CHAUMOND J. & WOLF T. (2020). DistilBERT, a distilled version of BERT : smaller, faster, cheaper and lighter.

SERETAN V. & WEHRLI E. (2007). Collocation translation based on sentence alignment and parsing. In (Benamara *et al.*, 2007), p. 401–410.

SIHAB-US-SAKIB S., RAHMAN M. R., FORHAD M. S. A. & AZIZ M. A. (2024). Cyberbullying detection of resource constrained language from social media using transformer-based approach. *Natural Language Processing Journal*, **9**, 100104. DOI : https ://doi.org/10.1016/j.nlp.2024.100104.

SOWLES S. J., MCLEARY M., OPTICAN A., CAHN E., KRAUSS M. J., FITZSIMMONS-CRAFT E. E., WILFLEY D. E. & CAVAZOS-REHG P. A. (2018). A content analysis of an online pro-eating disorder community on Reddit. *Body Image*, **24**, 137–144. DOI : 10.1016/j.bodyim.2018.01.001.

WANG W., WEI F., DONG L., BAO H., YANG N. & ZHOU M. (2020). MiniLM : Deep Self-Attention Distillation for Task-Agnostic Compression of Pre-Trained Transformers.