

# Détection d’anomalies textuelles par ensemble d’autoencodeurs robustes

Jeremie Pantin\*, Christophe Marsala\*

\*Sorbonne Université, CNRS, LIP6, Paris F-75005, France  
Prenom.Nom@lip6.fr,

**Résumé.** La détection d’anomalies en apprentissage automatique, un défi étudié dans divers domaines, suscite un intérêt croissant, présentant des défis uniques. Dans le cadre du texte, les approches existantes négligent souvent un aspect crucial: le type d’anomalie ciblé. De plus, les protocoles expérimentaux ne différencient pas les types d’anomalies textuelles, limitant la portée des travaux. Après avoir formalisé deux types d’anomalies (indépendantes et contextuelles), nous proposons AREnT une nouvelle approche utilisant des autoencodeurs ensemblistes robustes, connectés aléatoirement. Notre approche réussit à détecter des anomalies simples et complexes. On note également que AREnT excelle non seulement à fournir des résultats compétitifs par rapport aux méthodes existantes, mais également dans le traitement des anomalies contextuelles.

## 1 Introduction

La détection d’anomalies (Ruff et al., 2021) est une tâche récurrente consistant à estimer si une observation est considérée normale ou non, relativement à un ensemble de données. Une telle détection peut être effectuée selon plusieurs approches, par exemple à l’aide de modèles de classification unaire (Khan et Madden, 2014; Sohn et al., 2021). Souvent, trois principaux types d’anomalies sont mis en évidence : *indépendante*, *collective* et *contextuelle*.

Récemment, l’apparition de modèles de langage tels que GloVe (Pennington et al., 2014) et RoBERTa (Liu et al., 2019) ont offert des applications prometteuses pour l’étude du texte. Seules quelques approches se sont intéressées à la détection d’anomalies textuelles (Ruff et al., 2019; Manolache et al., 2021), s’appuyant généralement sur une représentation TF-IDF des textes (Schölkopf et al., 2001; Kannan et al., 2017). De même, les autoencodeurs (AE) qui sont grandement utilisés pour la détection d’anomalies avec des données de haute dimension (Chen et al., 2017; Kieu et al., 2019), et qui sont également efficaces sur d’autres types de données (Zhou et Paffenroth, 2017; Chen et al., 2018; Lai et al., 2020), sont pourtant peu utilisés dans le cadre d’anomalies textuelles. D’un autre côté, on peut aussi remarquer qu’il n’existe pas d’approches pour la détection d’anomalies textuelles en utilisant des méthodes ensemblistes.

Nous proposons une nouvelle approche se basant sur l’utilisation d’un ensemble qui se repose sur une représentation du texte par modèle de langage, d’AE connectés aléatoirement et une couche robuste par *Robust Subspace Recovery* (RSR) (Lerman et Maunu, 2018). Notre méthode, *AE Robustes Ensemblistes pour le Texte* (AREnT), met en œuvre un apprentissage de plusieurs sous-espaces visant à trouver des observations normales à partir de différents

## Détection d'anomalies textuelles par ensemble d'autoencodeurs robustes

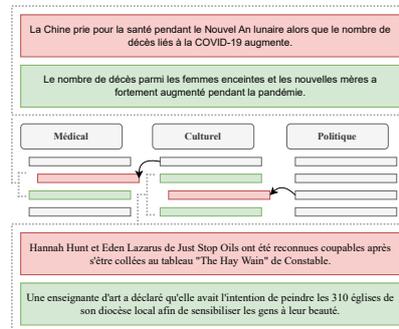


FIG. 1 – Illustration de différents scénarios de contamination.

*manifold*. L'utilisation d'un AE permet de projeter la distribution des données d'origine dans un sous-espace où les anomalies sont éparpillées (Lai et al., 2020).

Cet article se compose comme suit, les travaux existants sont décrits dans la Section 2, dans la Section 3 nous décrivons l'approche AREnT, dans la Section 4 nous présentons la définition des anomalies traitées dans ce travail, dans la Section 5 les expérimentations conduites sont présentées puis nous terminons par une conclusion et les perspectives de ce travail.

## 2 Travaux préliminaires

Afin d'évaluer les performances de AREnT face aux approches de l'état de l'art, une expérimentation à partir de 8 jeux de données est décrite. Pour cette expérimentation, on peut mettre en évidence un des problèmes dans le domaine de la détection d'anomalies textuelles qui réside dans la comparaison des approches proposées. En effet, pour évaluer une nouvelle approche, il est difficile de reproduire les protocoles expérimentaux et les résultats de l'état de l'art. Ainsi il y a une grande différence entre étudier des anomalies indépendantes et étudier des anomalies contextuelles (Mahapatra et al., 2012; Fouché et al., 2020). En présence d'anomalies indépendantes, un modèle de classification a besoin de faire la différence entre deux documents qui ont des sujets sans lien de parenté (par exemple, "sport" et "informatique"), tandis qu'en présence d'anomalies contextuelles, il faut identifier des documents qui ont un sujet parent en commun (eg. "tennis" et "football" qui ont "sport" comme sujet parent). La plupart des travaux de l'état de l'art tentent alors de détecter des anomalies sans distinguer ces 2 types, impliquant un problème d'expérimentation lors de la préparation des données (Kannan et al., 2017; Ruff et al., 2019; Lai et al., 2020). La Figure 1 illustre une telle distinction.

## 3 Autoencodeurs robustes ensemblistes pour les anomalies textuelles (AREnT)

**Autoencodeurs à connexions aléatoires** Soit  $X = \{x_1, \dots, x_N\}$  un corpus de  $N$  documents. Chaque document  $x \in X$  est décrit par un vecteur de dimension  $D$  :  $x = (x_1, \dots, x_D)$ . Un AE est un réseau de neurones composé de 2 éléments : un encodeur  $\mathcal{E}$  et un décodeur  $\mathcal{D}$ .

- l’encodeur  $\mathcal{E}$  prend en entrée la description  $(x_1, \dots, x_D)$  pour construire une représentation latente  $z = \mathcal{E}(x) \in \mathbb{R}^e$  de  $x$  dans l’espace  $\mathbb{R}^e$  de dimension  $e$ .
- le décodeur  $\mathcal{D}$  reconstruit à partir de  $z \in \mathbb{R}^e$  le document  $\hat{x} \in \mathbb{R}^D$  décrit dans l’espace d’origine : on a alors  $\hat{x} = \mathcal{D}(z)$ .

Pour AREnT, au lieu d’utiliser des AE entièrement connectés, nous proposons d’utiliser des AE connectés de manière aléatoire (entre 50% et 80% de connexions), de façon similaire à Chen et al. (2017).

**Couche robuste** Pour augmenter la robustesse de notre ensemble d’AE, nous utilisons l’approche de Lai et al. (2020). Dans cette approche, la couche RSR est une transformation linéaire  $\mathbf{A} \in \mathbb{R}^{d \times e}$  qui réduit la dimension à  $d$ . On note alors  $\hat{z}$  la représentation de  $z$  à travers la couche RSR, telle que  $\hat{z} = \mathbf{A}z^T \in \mathbb{R}^d$ . La matrice  $\mathbf{A}$  et les paramètres de  $\mathcal{E}$  et  $\mathcal{D}$  sont obtenus par la minimisation d’une fonction de perte.

Un autoencodeur RSR, noté  $(\mathcal{E}, \mathbf{A}, \mathcal{D})$ , vise à réduire de manière robuste et non linéaire la dimension des données originales (Lerman et Maunu, 2018). Un tel modèle représente les observations normales autour de leurs emplacements d’origine, et les anomalies éloignées de leurs emplacements d’origine. La fonction de perte en découlant est une agrégation de la fonction de perte de l’AE notée  $L_{AE}$  avec la fonction de perte de la couche RSR notée  $L_{RSR}$ .  $L_{AE}$  est basée sur la norme  $l_2$  et  $p > 0$  :

$$L_{AE}^p(\mathcal{E}, \mathbf{A}, \mathcal{D}) = \sum_{i=1}^N \|x_i - \hat{x}_i\|_2^p \quad (1)$$

$L_{RSR}$  comporte un premier terme qui impose à la couche RSR d’être robuste (estimation PCA) et un second terme qui impose à la projection d’être orthogonale :

$$L_{RSR}^q(\mathbf{A}) = \lambda_1 \sum_{i=1}^N \|z_i - \mathbf{A}^T \hat{z}_i\|_2^q + \lambda_2 \sum_{i=1}^N \|\mathbf{A}\mathbf{A}^T - \mathbf{I}_d\|_f^q \quad (2)$$

avec  $\mathbf{A}^T$  la transposée de  $\mathbf{A}$ ,  $\mathbf{I}_d$  de matrice  $d \times d$  et  $\|\cdot\|_f$  la norme de Frobenius.  $\lambda_1$  et  $\lambda_2$  sont des hyperparamètres et  $q = 1$  correspond à la norme optimale  $l_{p,q}$  (Maunu et al., 2019).

**Apprentissage d’ensembles et représentation du texte** L’idée principale derrière les méthodes d’ensemble (Ganaie et al., 2022) est qu’une combinaison de plusieurs modèles, également appelés *apprenants faibles*, a des résultats plus robuste que l’utilisation d’un unique modèle. Cette notion de robustesse peut s’observer face au compromis biais-variance ainsi qu’au problème de sur-apprentissage. Dans AREnT, nous utilisons l’erreur de reconstruction RSR de chaque AE, puis nous normalisons les scores de chaque détecteur de base via l’écart type. Nous prenons ensuite la valeur médiane pour chaque observation. Dans AREnT chaque apprenant faible est un autoencodeur tel que décrit dans la section précédente.

Dans ce travail nous proposons d’utiliser RoBERTa (Liu et al., 2019; Reimers et Gurevych, 2019) pour la représentation du texte contrairement à GloVe ou TFIDF qui sont utilisés par les approches de l’état de l’art. GloVe a été utilisée avec succès par Ruff et al. (2019); Manolache et al. (2021) dans leurs approches, en supplément de BERT.

## 4 Anomalies contextuelles

Un grand nombre de contributions ont traité les taxonomies d'anomalies/valeurs aberrantes (Zhang, 2013; Aggarwal, 2017; Ruff et al., 2021) et plusieurs types d'anomalies ont été proposées dans la littérature : anomalie indépendante, anomalie contextuelle et anomalie collective. Une taxonomie similaire peut être appliquée dans le domaine des données textuelles car différents types d'anomalies coexistent fréquemment au sein des documents d'un corpus donné.

Considérons un document étiqueté d'un corpus  $(x, y) \in X \times \mathcal{Y}$  et  $\zeta$  la catégorie normale de sous-ensemble  $X_\zeta \subseteq X$ . Nous définissons  $\mathcal{A}$  le sous-ensemble de toutes les anomalies telles que  $\mathcal{A} \subset X$ . Nous avons  $\mathcal{A} = X \setminus X_\zeta$ . En fonction de  $\mathcal{A}$ , nous faisons la distinction entre deux contraintes. On note  $P(y)$  le parent direct de  $y$  dans une hiérarchie d'étiquettes donnée. Premièrement, une observation  $x_i$  est considérée comme une anomalie si son sujet parent est différent du sujet de normalité  $\zeta$ , de telle sorte que :  $\mathcal{A}_p(\zeta) = \{P(\zeta) \neq P(y) | (a, y) \in \mathcal{A} \times Y\}$ . La seconde contrainte correspond aux documents qui n'appartiennent pas à  $X_\zeta$  mais partagent le même sujet parent à  $\zeta$ . Ces observations sont identifiées comme un second type d'anomalie : anomalie contextuelle. Ainsi, on note :  $\mathcal{A}_c(\zeta) = \{P(\zeta) = P(y) | (a, y) \in \mathcal{A} \times Y, \mathcal{A} \setminus \mathcal{A}_p\}$ .

## 5 Expérimentations

**Données** Récemment, des travaux récents (Kannan et al., 2017; Mahapatra et al., 2012; Ruff et al., 2019; Lai et al., 2020) utilisent des jeux de données de classification tels que Reuters-21578<sup>1</sup> et 20 Newsgroups<sup>2</sup> en effectuant une préparation artificielle à partir de leurs étiquettes. Pareillement, nous utilisons au total 8 jeux de données issus de la classification de texte, de l'analyse de sentiment et de détection de spam. Pour chaque sujet disponible, nous appliquons la préparation des anomalies indépendantes et contextuelles avec l'algorithme GenTO (Pantin et al., 2022). Pour des raisons d'équité entre chaque méthode et chaque corpus, nous fixons la taille du sous-ensemble de préparation à 350 et effectuons la moyenne des résultats sur 10 exécutions. Pour l'algorithme GenTO nous définissons le taux de contamination  $\nu = 0.1$ .

Pour 20 Newsgroups, nous séparons les thèmes sous sept sujets parents : computer, for-sale, motors, politics, religion, science, sports. Nous réorganisons les sujets de Reuters-21578 afin d'obtenir une hiérarchie basée sur les travaux de Toutanova et al. (2001) (nous enlevons les documents associés à plusieurs étiquettes). Ainsi, quatre sujets parents sont créés : commodities, finance, metals and energy. Pour DBpedia 14, nous créons la hiérarchie de sujets basée sur l'ontologie fournie<sup>3</sup>, comportant six sujets parents. Pour Web of Science, les sujets du troisième niveau sont répartis entre les parents correspondants du premier niveau. Ainsi, sept sujets parents sont présents, et pour les sujets enfants associés à plus d'un parent, nous conservons le plus grand ensemble d'enfants et supprimons les autres.

**Préparation** Nous utilisons GenTO pour préparer une contamination contextuelle sur tout sujet normal possible avec  $\nu = 0.1$  et une taille de 350. Tous les résultats sont réalisés en calculant les métriques d'évaluations AC (aire sous la courbe ROC) et AP (précision moyenne). Nous intégrons les résultats d'un AE à une classe et nous comparons également les résultats sur

1. <http://www.daviddlewis.com/resources/testcollections/reuters21578/>  
 2. <http://qwone.com/~jason/20Newsgroups/>  
 3. [mappings.dbpedia.org/server/ontology/classes/](http://mappings.dbpedia.org/server/ontology/classes/)

Contamination indépendante ( $\nu = 0.1$ )								
Model	Newsgroups	Reuters	WOS	DBpedia 14	Enron	SMS Spam	IMDB	SST2
OCSVM	0.948	0.917	0.981	0.993	0.723	0.693	0.539	0.575
PCC	<b>0.952</b>	0.938	0.982	0.992	0.724	0.685	0.542	0.576
OC-AE	0.697	0.732	0.856	0.837	0.592	0.514	0.517	0.499
RSRAE	0.949	0.940	0.982	0.994	0.731	0.704	0.540	0.577
CVDD	0.781	0.895	0.948	0.958	0.629	0.717	0.513	0.562
REA	0.884	0.704	0.935	0.918	0.636	0.553	0.665	0.614
AREnT	0.949	<b>0.953</b>	<b>0.989</b>	<b>0.991</b>	<b>0.749</b>	<b>0.898</b>	<b>0.704</b>	<b>0.627</b>
Contamination contextuelle ( $\nu = 0.1$ )								
Model	Newsgroups		Reuters		WOS		DBpedia 14	
	AP	AC	AP	AC	AP	AC	AP	AC
OCSVM	0.282	0.750	0.491	0.811	0.599	0.889	0.759	0.945
PCC	0.314	0.776	0.518	0.828	0.613	0.897	0.771	<b>0.954</b>
OC-AE	0.191	0.623	0.246	0.604	0.249	0.680	0.348	0.735
RSRAE	0.309	0.779	0.506	0.821	0.621	0.900	0.762	0.936
CVDD	0.131	0.613	0.481	0.824	0.623	0.916	0.646	0.919
REA	0.194	0.623	0.278	0.615	0.448	0.810	0.368	0.747
AREnT	<b>0.362</b>	<b>0.793</b>	<b>0.538</b>	<b>0.880</b>	<b>0.687</b>	<b>0.921</b>	<b>0.840</b>	0.951

TAB. 1 – Résultats des expérimentations conduites sur la contamination indépendante et contextuelle. La représentation du texte est RoBERTa.

un ensemble d’AE connectés de manière aléatoire (RAE) (Chen et al., 2017). L’architecture est similaire à Chen et al. (2017) et les AE suivent leurs paramètres. Il en va de même pour l’approche RSRAE qui suit la configuration de Lai et al. (2020). Nous maintenons également le nombre d’exécutions pour chaque contamination à 10.

Pour AREnT et RAE, la configuration est la même que l’AE et le nombre d’apprenants faibles est de 50. De plus, nous définissons également  $\lambda_1 = 0.1$  et  $\lambda_2 = 0, 1$  pour éviter les problèmes d’effondrement multiple et les solutions dégénérées. D’un autre côté, nous fixons le nombre d’époques à 50 et la probabilité de connexion aléatoire entre 50% et 80%.

Nous proposons de comparer notre approche avec OCSVM (Schölkopf et al., 2001), PCC (Shyu et al., 2003), un AE à une classe (OC-AE) et RSRAE (Lai et al., 2020). Pour OCSVM, PCC et OC-AE, nous utilisons l’implémentation de PyOD (Zhao et al., 2019) et pour RSRAE, nous utilisons leur implémentation. Nous présentons également les résultats de CVDD (Ruff et al., 2019) en utilisant la configuration recommandée dans leur article. Alternativement, nous proposons d’utiliser GenTO en considérant  $P(y) = P(\zeta)$  pour une contamination indépendante.

**Résultats** Nos résultats s’appuient sur trois points principaux : la contamination indépendante, la contamination contextuelle et la robustesse des scores du modèle. Le tableau 1 affiche les résultats de la contamination indépendante. Nous pouvons voir que AREnT surpasse tous les modèles sauf sur 20 Newsgroups. Si les résultats sont similaires, notre approche se démarque notamment des autres sur les corpus SMS Spam et IMDB. En conséquence, notre approche réussit à surclasser les autres méthodes aussi bien pour les corpus sémantiques, que pour les corpus de spam et de sentiments.

Le tableau 1 présente les résultats expérimentaux menés avec notre approche AREnT. Dans ces résultats, nous remarquons que AREnT surpasse les autres modèles avec la métrique AC et

## Détection d'anomalies textuelles par ensemble d'autoencodeurs robustes

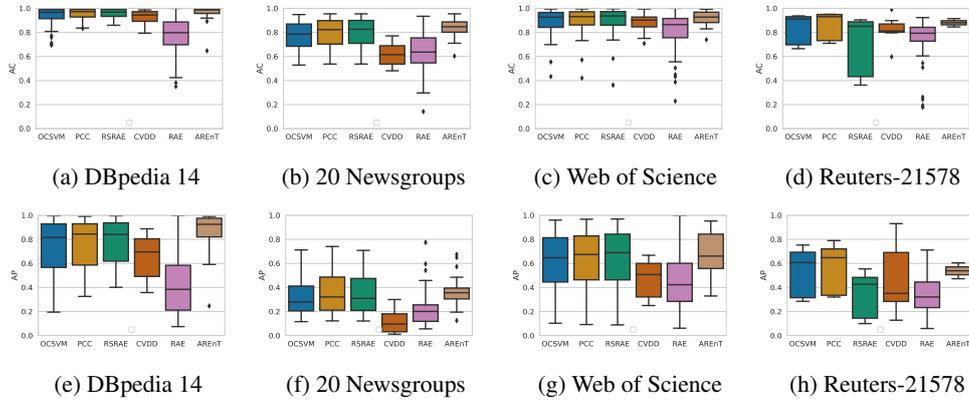


FIG. 2 – Résultats avec  $\nu = 0, 1$  et un nombre d'apprenants faibles de 25. En haut sont présentées les valeurs AC et en bas les valeurs AP.

la métrique AP. Nous pouvons voir que l'utilisation de AREnT permet d'atténuer la décision instable du RSRAE original. Également, nous constatons une différence de performances significative entre le corpus Web of Science et Reuters-21578. PCC est la seule approche qui réussit à battre notre approche sur la métrique AC de DBpedia 14. De plus, nous observons que l'AE unaire d'origine bénéficie grandement de la technique de connexion aléatoire et d'ensemble, car il comble l'écart avec les autres modèles.

Bien que nos performances soient compétitives, l'objectif principal de la détection d'anomalies avec des méthodes d'ensemble est d'atténuer le compromis biais-variance. La motivation de notre travail est d'obtenir des scores robustes contre des anomalies contextuelles. La Figure 2 affiche des résultats prometteurs de notre approche pour l'AC (Figures 2a,2b,2c et 2d) et l'AP (Figures 2e,2f,2g et 2h). Nous pouvons voir que la variance de notre modèle se démarque car la boîte est toujours plus petite que celle de ses concurrents. De plus, les scores minimum et maximum sont proches des scores médians, ce qui conclut que notre approche est plus efficace, plus robuste et peut utiliser un modèle de langage comme RoBERTA.

## 6 Conclusion

Dans ce travail, nous avons introduit AREnT, une approche d'ensemble d'AE robustes pour traiter les anomalies contextuelles dans le texte. Cette approche a démontré pouvoir obtenir des résultats significatifs et une grande robustesse face à deux types d'anomalies, avec une petite quantité de documents disponibles (350). Par ailleurs, nous avons mis en évidence plusieurs types d'anomalies coexistants au sein d'un corpus. Nos résultats montrent que les approches de la littérature mélangent des anomalies indépendantes et contextuelles et incorporent différentes difficultés de détection d'anomalies sans en avoir conscience. Une perspective de ces travaux consiste à étudier l'intégration des *attention heads* du modèle de langage RoBERTa afin de rendre la perception des éléments du texte plus précise. L'optimisation des manifolds dans l'ensemble est également une piste envisagée pour de futurs travaux.

## Références

- Aggarwal, C. C. (2017). Outlier Detection in Categorical, Text, and Mixed Attribute Data. In *Outlier Analysis*, pp. 249–272. Springer International Publishing.
- Chen, J., S. Sathe, C. Aggarwal, et D. Turaga (2017). Outlier detection with autoencoder ensembles. In *Proceedings of the 2017 SIAM international conference on data mining*, pp. 90–98. SIAM.
- Chen, Z., C. K. Yeo, B. S. Lee, et C. T. Lau (2018). Autoencoder-based network anomaly detection. In *2018 Wireless Telecommunications Symposium (WTS)*, pp. 1–5.
- Fouché, E., Y. Meng, F. Guo, H. Zhuang, K. Böhm, et J. Han (2020). Mining text outliers in document directories. In *2020 IEEE International Conference on Data Mining (ICDM)*, pp. 152–161. IEEE.
- Ganaie, M. A., M. Hu, A. Malik, M. Tanveer, et P. Suganthan (2022). Ensemble deep learning : A review. *Engineering Applications of Artificial Intelligence* 115, 105151.
- Kannan, R., H. Woo, C. C. Aggarwal, et H. Park (2017). Outlier Detection for Text Data. *SDM International Conference on Data Mining 17*, 489–497.
- Khan, S. S. et M. G. Madden (2014). One-class classification : taxonomy of study and review of techniques. *The Knowledge Engineering Review* 29(3), 345–374.
- Kieu, T., B. Yang, C. Guo, et C. S. Jensen (2019). Outlier detection for time series with recurrent autoencoder ensembles. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pp. 2725–2732. International Joint Conferences on Artificial Intelligence Organization.
- Lai, C.-H., D. Zou, et G. Lerman (2020). Robust subspace recovery layer for unsupervised anomaly detection. In *International Conference on Learning Representations*.
- Lerman, G. et T. Maunu (2018). An overview of robust subspace recovery. *Proceedings of the IEEE* 106(8), 1380–1410.
- Liu, Y., M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, et V. Stoyanov (2019). Roberta : A robustly optimized bert pretraining approach.
- Mahapatra, A., N. Srivastava, et J. Srivastava (2012). Contextual Anomaly Detection in Text Data. *Algorithms* 5(4), 469–489.
- Manolache, A., F. Brad, et E. Burceanu (2021). DATE : Detecting anomalies in text via self-supervision of transformers. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics : Human Language Technologies*, pp. 267–277. Association for Computational Linguistics.
- Maunu, T., T. Zhang, et G. Lerman (2019). A well-tempered landscape for non-convex robust subspace recovery. *Journal of Machine Learning Research* 20(37), 1–59.
- Pantin, J., C. Marsala, et M.-J. Lesot (2022). Analyse de données aberrantes pour le texte : Taxonomie et étude expérimentale. In *Extraction et Gestion des Connaissances-EGC'2022*, pp. 15–26.
- Pennington, J., R. Socher, et C. Manning (2014). GloVe : Global Vectors for Word Representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543. Association for Computational Linguistics.

- Reimers, N. et I. Gurevych (2019). Sentence-bert : Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics.
- Ruff, L., J. R. Kauffmann, R. A. Vandermeulen, G. Montavon, W. Samek, M. Kloft, T. G. Dietterich, et K.-R. Müller (2021). A Unifying Review of Deep and Shallow Anomaly Detection. *Proceedings of the IEEE* 109(5), 756–795.
- Ruff, L., Y. Zemlyanskiy, R. Vandermeulen, T. Schnake, et M. Kloft (2019). Self-Attentive, Multi-Context One-Class Classification for Unsupervised Anomaly Detection on Text. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 4061–4071. Association for Computational Linguistics.
- Schölkopf, B., J. C. Platt, J. Shawe-Taylor, A. J. Smola, et R. C. Williamson (2001). Estimating the support of a high-dimensional distribution. *Neural Computation* 13(7), 1443–1471.
- Shyu, M.-L., S.-C. Chen, K. Sarinapakorn, et L. Chang (2003). A novel anomaly detection scheme based on principal component classifier. Technical report, Miami University.
- Sohn, K., C.-L. Li, J. Yoon, M. Jin, et T. Pfister (2021). Learning and evaluating representations for deep one-class classification. In *International Conference on Learning Representations*, Volume 9.
- Toutanova, K., F. Chen, K. Popat, et T. Hofmann (2001). Text classification in a hierarchical mixture model for small training sets. In *Proceedings of the tenth international conference on Information and knowledge management*, pp. 105–113.
- Zhang, J. (2013). Advancements of outlier detection : A survey. *ICST Transactions on Scalable Information Systems* 13(1), 1–26.
- Zhao, Y., Z. Nasrullah, et Z. Li (2019). Pyod : A python toolbox for scalable outlier detection. *Journal of Machine Learning Research* 20(96), 1–7.
- Zhou, C. et R. C. Paffenroth (2017). Anomaly detection with robust deep autoencoders. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '17*, pp. 665–674. Association for Computing Machinery.

## Summary

Anomaly detection in machine learning is a recurring challenge studied in various domains, which has a blossoming interest and present unique challenges. One issue of existing approaches is that they overlook a crucial aspect: the type of targeted anomaly. Furthermore, experimental protocols often do not differentiate existence of several textual anomalies, limiting the scope of the work. After formalizing two types of anomalies (independent and contextual), we propose a novel approach using robust ensemble autoencoders, which are randomly connected. Our approach successfully detects both simple and complex anomalies. It is noteworthy that our approach excels not only in providing competitive results compared to existing methods but also in handling contextual anomalies.