Text Representation through Multimodal Variational Autoencoder for One-Class Learning

Marcos Paulo Silva Gôlo
marcosgolo@usp.br
Institute of Mathematics and Computer Sciences, USP
São Carlos, São Paulo

Ricardo Marcondes Marcacini
ricardo.marcacini@icmc.usp.br
Institute of Mathematics and Computer Sciences, USP
São Carlos, São Paulo

Abstract

Multi-class learning (MCL) methods perform Automatic Text Classification (ATC), which requires labeling for all classes. MCL fails when there is no well-defined class information and requires a great effort in labeling. One-Class Learning (OCL) can mitigate these limitations since the training only has instances from one class, reducing the labeling effort and making the ATC more appropriate for open-domain applications. However, OCL is more challenging due to the lack of counterexamples. Even so, most studies use unimodal representations, even though different domains contain other information (modalities). Thus, this study proposes the Multimodal Variational Autoencoder (MVAE) for OCL. MVAE is a multimodal method that learns a new representation from more than one modality, capturing the characteristics of the interest class in an adequate way. MVAE explores semantic, density, linguistic, and spatial information modalities. The main contribution is a new multimodal method for representation learning on OCL scenarios considering few instances to train with state-of-the-art results in three domains.

Keywords: Text Classification, One-Class Learning, Multimodal Variational Autoencoder

1 Introduction

ATC assigns a previously defined label in unlabeled textual documents. MCL is a strategy for ATC. In MCL, the user must know all classes of the problem and label documents for all those classes in the training step, implying two limitations: when the user does not label examples for all domain classes and when a new domain class comes up. Thus, MCL requires a greater effort to label the instances for each class [1].

One approach that mitigates some limitations presented by the MCL is One-Class Learning (OCL) [2, 22]. OCL uses only instances from one class to learn [22]. OCL will be able to identify whether an instance belongs to the interest class, reducing the labeling effort and being more appropriate for open-domain applications or one-class applications [6, 10].

The learning process is more challenging for the OCL due to the lack of counterexamples. Even so, studies use the traditional Bag-of-Words (BoW) technique [12, 16]. Other studies explore dimensionality reduction techniques [8, 14]. Finally, studies used language models via neural networks [17, 19]. This study highlights that the related work generates a representation focused on the text sentences. However, domains contain other information useful for learning, such as topics, geographic, and semantic information [11, 25].

Despite the benefits of representations generated through multimodal learning [11, 15], using multimodality to represent texts in the OCL scenario is a gap in the literature. Thus, this study has two research challenges related to multimodal learning. First, obtain multimodal representations suitable for OCL in ATC. Second, obtain these representations with few labeled instances. These research challenges involve investigating appropriate representations to reduce dependence on large interest training sets, outperforming other representations commonly used in the OCL literature.

Given the research challenges, this study proposed the Multimodal Variational Autoencoder (MVAE) for OCL. MVAE is a representation learning method that learns a representation from multiple modalities through a neural network based on VAEs that are generative models considered one of the state-of-the-art for text representation learning [23, 24]. Moreover, another research goal is to analyze the proposal in real-world applications with few labeled instances, such as detecting fake news, relevant reviews, and events, to verify the generalizability in different domains.

This study has two main contributions. First, **MVAE for OCL in ATC**: the study proposes an MVAE architecture that generates more suitable textual representations for OCL. MVAE explores as modalities: (i) embeddings from the BERT multilingual (semantic knowledge); (ii) density information from the high-density regions; (iii) linguistic structure of the texts; and (iv) geolocation data (latitude and longitude). Second, **MVAE in three real scenarios**. The study detects fake news [6, 7] and relevant app reviews [5] with MVAE representations via BERT and density modalities. This study also proposes MVAE for representing events considering BERT, geolocation, and density modalities [9, 10]. MVAE proved your robustness for scenarios with few labeled instances in the three domains, reducing the labeling effort and outperforming other representations methods.

In: V Concurso de Teses e Dissertações (CTD 2023), Ribeirão Preto, Brasil. Anais Estendidos do Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia). Porto Alegre: Sociedade Brasileira de Computação, 2023.

ISSN 2596-1683

2 Multimodal Variational Autoencoder

MVAE is able to explore different modalities. This study explores two main modalities: the DistilBERT [3] representation and a proposed modality for density representation. The proposed density modality is a visual modality based on the different topics of the interest class. Specifically, our proposal explores clustering methods to identify high-density regions from texts (topics). Then, the study extracts features from each cluster through statistical measures that describe the merits of the cluster structure.

Consider a clustering with k clusters, i.e., $\mathcal{D} = C_1 \cup C_2, \cup \cdots \cup C_k$, in which C_j is a cluster of documents, and $2 \leq k < m$. Then, the study applies the silhouette coefficient [18] in order to measure if a document belongs to a single topic or contains mixed topics. The silhouette for a document d_i represented by the embeddings of BERT $\boldsymbol{\lambda}_i$ assigned to a cluster C_j is given by Equation 1. The silhouette values range from -1 to +1. A high value indicates that a document is well-matched to its cluster and weakly matched to other clusters. We represent the density information by concatenating silhouette values considering each document in different clustering settings. Given u different clustering settings (different ks), Equation 2 shows Density modality.

$$s(\boldsymbol{d}_i, k) = \frac{\beta(\boldsymbol{\lambda}_i) - \alpha(\boldsymbol{\lambda}_i)}{\max(\alpha(\boldsymbol{\lambda}_i), \beta(\boldsymbol{\lambda}_i))}, \quad (1) \qquad \boldsymbol{\delta}_i = \{s(\boldsymbol{d}_i, k_1), s(\boldsymbol{d}_i, k_2), .. ., s(\boldsymbol{d}_i, k_u - 1), s(\boldsymbol{d}_i, k_u)\}, \quad (2)$$

in which $\alpha(\boldsymbol{\lambda}_i)$ is the average distance of $\boldsymbol{\lambda}_i$ to all documents of cluster C_j, $\beta(\boldsymbol{\lambda}_i)$ is the average distance of a document $\boldsymbol{\lambda}_i$ to all documents of the closest cluster C_o, $o \neq j$, and $s(\boldsymbol{d}_i, k_j)$ is the silhouette of $\boldsymbol{d}_i$ in cluster setting with k_j clusters.

After obtaining the modalities in a structured way, the study must fuse them. In multimodal learning, early fusion is one of the most common types of fusion. The early fusion can be represented by combining modalities before the machine learning process. For instance, early fusion can be done using simple operators such as concatenation, [13]. An advantage of early fusion is using only one data representation in the learning process. However, early fusion has some challenges, such as dealing with different dimensions and importance levels of each modality. To deal with these challenges, this study uses early fusion operators based on neural networks [4]. Therefore, MVAE uses dense layers with the same shape as the first layers, and in the second layer, MVAE uses a merging layer that works as an early fusion operator. Thus, the study can choose different operators for our early fusion and still use modalities with different sizes while our neural network learns the importance of each modality.

MVAE is a neural network with an encoder and decoder step, such as an Autoencoder. However, our MVAE is a Variational Autoencoder (VAE) variant. Thus, MVAE has useful properties for representation learning [23]. The MVAE performs a sampling step based on a previous distribution model to generate a bottleneck. Therefore, MVAE bias the learning through a prior informed distribution model, which is attractive for OCL because the representations generated by MVAE will preserve the main characteristics from the interest class representations and the model distribution, generating a region of the interest class [23, 24].

For our modalities $\boldsymbol{\lambda}_i$ and $\boldsymbol{\delta}_i$, MVAE assumes that z_i generates $\boldsymbol{\lambda}_i$ and $\boldsymbol{\delta}_i$ using $p(z_i|\boldsymbol{\lambda}_i, \boldsymbol{\delta}_i) = \frac{p(\boldsymbol{\lambda}_i, \boldsymbol{\delta}_i|z_i)p(z_i)}{p(\boldsymbol{\lambda}_i, \boldsymbol{\delta}_i)}$, in which $p(\boldsymbol{\lambda}_i, \boldsymbol{\delta}_i)$ is defined via $p(\boldsymbol{\lambda}_i, \boldsymbol{\delta}_i) = \int p(\boldsymbol{\lambda}_i, \boldsymbol{\delta}_i|z_i)p(z_i)dz$. Integrals are computationally intractable. Thus, MVAE uses variational inference, an approximation technique, to solve the limitation. Therefore, MVAE approximates $p(z_i|\boldsymbol{\lambda}_i, \boldsymbol{\delta}_i)$ to $q(z_i|\boldsymbol{\lambda}_i, \boldsymbol{\delta}_i)$ (treatable distribution) through the Kullback-Leibler (KL) divergence. Finally, MVAE optimizes the marginal likelihood ($p(\boldsymbol{\lambda}_i, \boldsymbol{\delta}_i)$) using the log of the marginal likelihood by Equation 3.

$$\log p_\Theta(\boldsymbol{\lambda}_i, \boldsymbol{\delta}_i) = KL(q_\Phi(z_i|\boldsymbol{\lambda}_i, \boldsymbol{\delta}_i)||p_\Theta(z_i|\boldsymbol{\lambda}_i, \boldsymbol{\delta}_i)) + \mathcal{L}(\Theta, \Phi; \boldsymbol{\lambda}_i, \boldsymbol{\delta}_i), \quad (3)$$

$$\mathcal{L}(\Theta, \Phi; \boldsymbol{\lambda}_i, \boldsymbol{\delta}_i) = \mathbb{E}_{q_\Phi(z_i|\lambda_i, \delta_i)} \log p_\Theta(\boldsymbol{\lambda}_i, \boldsymbol{\delta}_i|z_i) - KL(q_\Phi(z_i|\boldsymbol{\lambda}_i, \boldsymbol{\delta}_i)||p_\Theta(z_i)). \quad (4)$$

MVAE minimizes the first term from Equation 3 maximizing the second term (Equation 4), in which the first term is the neural network reconstruction loss and the second is the KL loss from $q_\Phi(z_i|\boldsymbol{\lambda}_i, \boldsymbol{\delta}_i)$ and $p_\Theta(z_i)$ (prior knowledge from distributed model). This study replaces the term $p_\Theta(z_i)$ with the $\mathcal{N}(z_i; 0, 1)$.

This study uses OCL algorithms to detect texts of interest. In OCL, the user defines an interesting class, and the OCL algorithm learns a classification model considering only documents of the interest class. Thus, the algorithm classifies a new document belonging to the interest class or not. Recent studies indicate that OCL is a competitive classification strategy with the advantage of reducing the user labeling effort [2, 22].

Among the OCL algorithms in the literature, the study chose the Support Vector Data Description (SVDD) [21] since it achieves good performance when the user represents the instances appropriately, is considered one of the state-of-the-art in the area of OCL and use in different application and scenarios [2]. The SVDD from [21] aims to generate the smallest hypersphere (given a radius and center) that involves the examples of the interest class. Examples allocated to the edge of the hypersphere are the support vectors. Thus, the SVDD wants to find the minimum volume hypersphere involving the interest documents according to Equation 5.

$$\boldsymbol{\mu}_{(c)} = \underset{\boldsymbol{\mu} \in U}{\arg\min} \max_{1 \leq i \leq m} ||\varphi(\boldsymbol{d}_i) - \boldsymbol{\mu}||^2, \quad (5)$$

in which $\boldsymbol{\mu} \in U$ is a possible center in the feature space U associated with the function kernel φ, $\varphi(\boldsymbol{d}_i)$ maps $\boldsymbol{d}_i$ into a feature space defined according to the kernel, and $\boldsymbol{\mu}_{(c)}$ is the hypersphere center in which the highest distance between $\varphi(\boldsymbol{d}_i)$ to $\boldsymbol{\mu}_{(c)}$ is minimal.

SVDD classifies a document as belonging to the interest class if its distance from the center is less than the radius r of the hypersphere. SVDD minimizes the function in Equation 6 subeject to Equation 7, in which ε_{d_i} is the external distance between $\varphi(d_i)$ and the surface of the hypersphere and $\nu \in (0, 1]$ defines the smoothness level of the hypersphere volume.

$$\min_{\mu, \varphi, r} \quad r^2 + \frac{1}{m} \sum_{i=1}^{m} \frac{\varepsilon_{\boldsymbol{d}_i}}{\nu}, \quad (6) \qquad ||\varphi(\boldsymbol{d}_i) - \boldsymbol{\mu}_{(c)}||^2 \leq r^2 + \varepsilon_{\boldsymbol{d}_i} \quad \forall i = 1, ..., m. \quad (7)$$

3 Results and Discussion

This study generated results published in five articles that can be grouped into 3 groups of interest text detection: (i) fake news [6, 7]; (ii) relevant app reviews [5]; and (iii) events of interest [9, 10]. This study presents the results according to the three groups. In all experimental evaluations, the studies propose to compare the MVAE with other unimodal and multimodal state-of-the-art representation methods. The study evaluates the representations in the one-class learning scenario with the SVDD. Our goal is to demonstrate that the representations generated by MVAE outperform others commonly used in the literature for fake news, relevant app reviews, and interest event classification to show the robustness of the method with little data to train in different domains.

The fake news study [6] uses three fake news datasets. Furthermore, in order to evaluate the approach, this work proposes an adaptation of the procedure k-Fold Cross-Validation, considering the OCL classification scenario with less labeling, in which the study uses 30%, 50%, 70%, and 100% of one fold to train (3, 5, 7, and 10% of the fake news) and nine folds to test (more details in dissertation). The relevant reviews study [5] uses three datasets created by [20] in the experimental evaluation and the representation proposed by them, which the study calls Maalej. The study uses the same adaptation of procedure k-Fold Cross-Validation mentioned above. The interest event study [10] uses 10 event collections for the experimental evaluation. The study uses event dates to separate training and testing. Events with older dates are from the training set. The study explored using a few labeled instances in the training set. Thus, the study explored using 60 and 2000 events in the training set. In the test, the study uses 4000 interest events. Also, the study randomly selected 4000 events from different event datasets and added them to the test set.

Table 1 presents the results for the three textual collections considering each percentage of fake news used in training. The results compare the ten representation models of fake news. The proposed MVAE obtained the highest F_1-Score in ten of the twelve evaluated scenarios. In the remaining two scenarios, the density information got the highest F_1. BoW obtained the lowest F_1 in all scenarios. Table 2 presents the highest values of F_1-Score obtained by ten app review representation techniques on the ARE app review collection. Bold values indicate that the method obtained the highest value in the column. Table 3 presents an analysis in relation to the Triple-VAE using 60 events compared to the other methods using 2000 to train. Triple-VAE achieved a higher F_1-Score than all other methods in 7 of 10 datasets. Furthermore, Triple-VAE achieved a higher F_1 in 3 collections than all other methods except DBERTML. Thus, considering the scenario with fewer labeled events closer to real-world applications, Triple-VAE was the best method for event detection as it achieved the best F_1-Scores.

For Table 1, the MVAE trained with only 3% of labeled fake news got better F_1-Scores than the other methods when these consider 10% of labeled fake news for most scenarios. Density representation got the best results considering 7% and 10% in the Fake Br. More qualitative results, such as a comparison of 2D projections, are found in the dissertation. We extend this study in the [7] article. Furthermore, for Table 2, MVAE obtained better the highest F_1-Scores than the other methods. Moreover, MVAE with only 3% relevant reviews obtained the highest F_1-Scores than the other methods with 10%. Finally, in Table 3, TripleVAE, with 60 events,

Table 1. Higher values of F_1-Scores of each representation considering the training percentage in the three datasets (DT). DBERTML is the DistilBert Multilingual model. VAE is a unimodal model with the DBERTML as input.

DT	%	BoW	DBERTML	Density	VAE	MVAE
Fake Br	3%	0.600±0.01	0.574±0.01	0.621±0.03	0.632±0.01	**0.642±0.00**
	5%	0.603±0.01	0.602±0.01	0.633±0.02	0.637±0.00	**0.644±0.00**
	7%	0.607±0.00	0.618±0.01	**0.647±0.02**	0.637±0.00	0.645±0.00
	10%	0.610±0.00	0.628±0.01	**0.650±0.01**	0.638±0.00	0.646±0.00
FCN	3%	0,556±0.02	0.426±0.06	0.487±0.07	0.741±0.04	**0.805±0.02**
	5%	0,582±0.01	0.568±0.05	0.575±0.08	0.796±0.03	**0.813±0.03**
	7%	0,591±0.01	0.640±0.03	0.584±0.06	0.801±0.01	**0.811±0.02**
	10%	0,596±0.01	0.706±0.02	0.625±0.03	0.804±0.02	**0.808±0.02**
FNN	3%	0,327±0.01	0.321±0.01	0.325±0.02	0.365±0.01	**0.395±0.03**
	5%	0,344±0.01	0.345±0.01	0.345±0.04	0.367±0.01	**0.403±0.03**
	7%	0,349±0.00	0.353±0.01	0.353±0.01	0.367±0.00	**0.397±0.01**
	10%	0,353±0.00	0.363±0.01	0.357±0.02	0.367±0.00	**0.393±0.01**

Table 2. Highest F_1-Scores from SVDD for each representation technique on the **ARE** dataset. MAE is a multimodal autoencoder. Other methods are the same present in Table 1.

Methods	3%	5%	7%	10%
BoW	0.55±0.02	0.57±0.02	0.58±0.01	0.60±0.01
Maalej	0.66±0.02	0.67±0.01	0.67±0.01	0.67±0.01
DBERTML	0.67±0.02	0.68±0.02	0.68±0.01	0.67±0.01
Density	0.66±0.03	0.66±0.02	0.65±0.02	0.64±0.02
VAE	0.68±0.01	0.69±0.02	0.69±0.01	0.67±0.01
MAE	0.70±0.01	0.71±0.01	0.72±0.02	0.72±0.01
MVAE	**0.72±0.03**	**0.75±0.02**	**0.74±0.03**	**0.74±0.01**

Table 3. Results in ten datasets considering the $\boldsymbol{F_1}$. Lat-Long is the latitude and longitude concatenated. Concat is the concatenated representation of the modalities. AE and VAE are multimodal methods trained in the Concat representation. TripleVAE is the MVAE method with three modalities.

Datasets	Unimodal			Trimodal			
	DBERTML	Lat-Long	Density	Concat	AE	VAE	TripleVAE
War	**0,855**	0,665	0,732	0,681	0,688	0,689	0,780
Tsunami	**0,933**	0,647	0,680	0,670	0,685	0,680	0,918
Covid	**0,955**	0,677	0,748	0,737	0,773	0,775	0,946
Corruption	0,931	0,676	0,665	0,692	0,692	0,691	**0,958**
Earthquake	0,912	0,660	0,693	0,671	0,667	0,679	**0,916**
Immigration	0,928	0,668	0,664	0,693	0,781	0,814	**0,950**
Racism	0,940	0,666	0,825	0,694	0,754	0,786	**0,964**
Inflation	0,950	0,666	0,796	0,681	0,659	0,662	**0,953**
Terrorism	0,925	0,676	0,690	0,683	0,681	0,682	**0,937**
Agriculture	0,914	0,657	0,677	0,670	0,728	0,728	**0,979**

outperforms other methods with 2000 events in 7 of 10 scenarios. Generally, MVAE was better than the other methods in learning highly non-linear relationships, redundancies, and dependencies between modalities, structuring the texts in a dimensional space more suitable for SVDD. Thus, our proposal structures texts with more representativeness of their modalities in relation to the other methods. All studies performed Friedman's statistical test and Nemenyi's post-test to compare the methods. In addition to obtaining the best average ranking, MVAE got statistical differences from unimodal methods. The tests are in the dissertation.

4 Conclusions and Future Work

This study presents a multimodal method developed to represent textual data considering the scenario of ATC through OCL. The method allows the use of a different number of modalities with different dimensions. The study explores three real-world application domains: (i) fake news classification; (ii) relevant app reviews detection; and (iii) web sensing. The study carried out an extensive empirical evaluation considering several MVAE architectures, textual languages, and different sizes of training sets. The study highlights the following innovations and contributions:

Multimodal method to represent the texts in OCL for ATC: the manuscript author proposes and develops a new multimodal method called Multimodal Variational Autoencoder (MVAE) and explores as modalities: (i) DistilBERT multilingual embeddings; (ii) density information; (iii) linguistic features; and (iv) geolocation. Any study can extend MVAE to use more than three modalities.

Detecting fake news, relevant reviews, and interest events: MVAE represents texts for classification through OCL. The study explores three modalities for fake news: DistilBERT, density, and linguistic features; two modalities for relevant reviews: DistilBERT and density; and three modalities for interest event: DistilBERT, density, and geolocation. The study highlights satisfactory performance, outperforming other state-of-the-art methods.

Textual collection involving news events for web sensing tasks: In the study [9], the study collects 183 textual datasets for the OCL. Each textual dataset has 6000 texts from the event titles of the Global Data of Events, Language, and Tone project. The study creates an OCTCMG library in the public repository https://github.com/GoloMarcos/OCTCMG.

Source code of the proposed MVAEs for the different applications: all source codes developed are available to the community at https://github.com/GoloMarcos/.

As the main Limitations, the study proposes a supervised method for ATC through OCL. Furthermore, MVAE needs a one-class algorithm, i.e., the study has two separate steps: text representation and classification. In this sense, future work is to investigate semi-supervised learning for OCL. Furthermore, other future work is to classify and represent texts through OCL in an end-to-end way, i.e., with a single learning process.

References

[1] Charu Aggarwal. 2018. *Machine Learning for Text*. Springer Publishing Company, online.

[2] Shamshe Alam, Sanjay Kumar Sonbhadra, Sonali Agarwal, and P Nagabhushan. 2020. One-class support vector classifiers: A survey. *Knowledge-Based Systems* 196 (2020), 1–19.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proc. of the 2019 Conf. of the North American Chapter of the ACL*. ACL, Minnesota, 4171–4186.

[4] Jing Gao, Peng Li, Zhikui Chen, and Jianing Zhang. 2020. A survey on deep learning for multimodal data fusion. *Neural Computation* 32, 5 (2020), 829–864.

[5] Marcos Gôlo, Adailton Araújo, Rafael Rossi, and Ricardo Marcacini. 2022. Detecting relevant app reviews for software evolution and maintenance through multimodal one-class learning. *Information and Software Technology* 151 (2022), 106998.

[6] Marcos Gôlo, Mariana Caravanti, Rafael Rossi, Solange Rezende, Bruno Nogueira, and Ricardo Marcacini. 2021. Learning Textual Representations from Multiple Modalities to Detect Fake News Through One-Class Learning. In *Proc. of the Brazilian Symposium on Multimedia and the Web*. ACM, online, 197–204.

[7] Marcos Gôlo, Mariana de Souza, Rafael Rossi, Solange Rezende, Bruno Nogueira, and Ricardo Marcacini. 2023. One-class learning for fake news detection through multimodal variational autoencoders. *Engineering Applications of Artificial Intelligence* 122 (2023), 106088.

[8] Marcos Gôlo, Ricardo Marcacini, and Rafael Rossi. 2019. An extensive empirical evaluation of preprocessing techniques and supervised one class learning algorithms for text classification. In *Proc. of the National Meeting on Artificial and Computational Intelligence*. SBC, Salvador, Brazil, 262–273.

[9] Marcos Gôlo, Rafael Rossi, and Ricardo Marcacini. 2021. Learning to sense from events via semantic variational autoencoder. *Plos one* 16, 12 (2021), e0260701.

[10] Marcos Gôlo, Rafael Rossi, and Ricardo Marcacini. 2021. Triple-VAE: A Triple Variational Autoencoder to Represent Events in One-Class Event Detection. In *Proceeding of the 2021 National Meeting on Artificial and Computational Intelligence*. SBC, online, 643–654.

[11] Wenzhong Guo, Jianwen Wang, and Shiping Wang. 2019. Deep multimodal representation learning: A survey. *IEEE Access* 7 (2019), 63373–63394.

[12] D. Junior and R. Rossi. 2017. Classificaçao automática de textos utilizando aprendizado supervisionado baseado em uma unica classe. *TCC em Sistemas de Informação* 1, 1 (2017), 24 pages.

[13] Aggelos K Katsaggelos, Sara Bahaadini, and Rafael Molina. 2015. Audiovisual fusion: Challenges and new approaches. *IEEE* 103, 9 (2015), 1635–1653.

[14] B Shravan Kumar and Vadlamani Ravi. 2017. Text Document Classification with PCA and One-Class SVM. In *Int. Conf. on Frontiers in Intel. Computing: Theory and Applications*. Springer, Odisa, India, 107–115.

[15] Yingming Li, Ming Yang, and Zhongfei Zhang. 2018. A survey of multi-view representation learning. *IEEE transactions on knowledge and data engineering* 31, 10 (2018), 1863–1883.

[16] Larry Manevitz and Malik Yousef. 2007. One-class document classification via neural networks. *Neurocomputing* 70, 7-9 (2007), 1466–1481.

[17] Hemanth Kumar Reddy Mayaluru. 2020. *One Class Text Classification using an Ensemble of Classifiers*. Ph. D. Dissertation. Rheinische Friedrich-Wilhelms-Universität Bonn.

[18] Peter Rousseeuw. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 20 (1987), 53–65.

[19] Lukas Ruff, Yury Zemlyanskiy, Robert Vandermeulen, Thomas Schnake, and Marius Kloft. 2019. Self-attentive, multi-context one-class classification for unsupervised anomaly detection on text. In *Proc. of the Meeting of the Association for Computational Linguistics*. ACL, Florence, Italy, 4061–4071.

[20] Christoph Stanik, Marlo Haering, and Walid Maalej. 2019. Classifying multilingual user feedback using traditional machine learning and deep learning. In *Int. Conf. Requirements Engineering*. IEEE, Jeju, Korea, 220–226.

[21] David MJ Tax and Robert PW Duin. 2004. Support vector data description. *Machine learning* 54 (2004), 45–66.

[22] David Martinus Johannes Tax. 2001. *One-class classification: Concept learning in the absence of counter-examples*. Ph. D. Dissertation. Technische Universiteit Delft.

[23] Hongzhi Wang, Mohamed Jaward Bah, and Mohamed Hammad. 2019. Progress in outlier detection techniques: A survey. *IEEE Access* 7 (2019), 107964–108000.

[24] Jiacheng Xu and Greg Durrett. 2018. Spherical Latent Spaces for Stable Variational Autoencoders. In *Proc. of the Conf. on Empirical Methods in NLP*. ACL, Brussels, Belgium, 4503–4513.

[25] Han Zhou, Hongpeng Yin, Hengyi Zheng, and Yanxia Li. 2020. A survey on multi-modal social event detection. *Knowledge-Based Systems* 195 (2020), 105695.