Knowledge-Based Systems xxx (xxxx) xxx

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Gated Mixture Variational Autoencoders for Value Added Tax audit case selection☆

Christos Kleanthous [a,b], Sotirios Chatzis [a,*]

[a] Department of Electrical Eng., Computer Eng., and Informatics Cyprus University of Technology, Limassol 3036, Cyprus
[b] Cyprus Tax Department, Nicosia, Cyprus

ARTICLE INFO

Article history:
Received 20 May 2019
Received in revised form 14 September 2019
Accepted 17 September 2019
Available online xxxx

Keywords:
Value Added Tax
Audit selection
Variational autoencoder
Finite mixture model

ABSTRACT

In this work, we address the problem of targeted Value Added Tax (VAT) audit case selection by means of machine learning. This is a challenging problem that has remained rather elusive for EU-based Tax Departments, due to the inadequate quantity of tax audits that can be used for conventional supervised model training. To this end, we devise a novel Gated Mixture Variational Autoencoder deep network, that can be effectively trained with data from a limited number of audited taxpayers, combined with a large corpus of filed VAT returns. This gives rise to a semi-supervised learning framework that leverages the latest advances in deep learning and robust regularization using variational inference. We developed our approach in collaboration with the Cyprus Tax Department and experimentally deployed it to facilitate its audit selection process; to this end, we used actual VAT data from Cyprus-based taxpayers. This way, we obtained strong empirical evidence that our approach can greatly facilitate the VAT audit case selection process. Specifically, we obtained up to 76% out-of-sample accuracy in detecting whether a significant tax yield will be generated from a specific prospective VAT audit.

1. Introduction

Valued Added Tax (VAT) is a consumption tax charged on the value of almost all the goods and services sold or consumed within the European Union (EU). It constitutes an indirect tax collected by enterprises on behalf of the state, and is ultimately paid by the final consumer. As such, it represents an important source of revenue for all EU Member States; according to the European Commission Taxation Trends Report, 2018 edition [1], indirect taxes comprise more than 30% of the total tax revenue in the EU.

The European commission uses the concept of VAT-gap to estimate the lost revenue of the EU Member States due to taxpayer non-compliance with the VAT legislation. It is defined as the difference between the estimated VAT amount that should have been collected and the actually collected amount. The latest European Commission study [2] estimated the VAT-gap at 12.3% (147 billion Euros) of the total expected VAT revenue in the EU; in Cyprus it stands at around 5% (83 million Euros). This exemplifies how important it is that we come up with effective methods for reducing the VAT-gap.

Tax administration is the responsibility of each EU member state tax authority. The primary goal is to collect the taxes due, according to the local tax legislation, in a sustainable manner. Ignorance, reckless behavior, tax evasion, and tax system inefficiencies are major sources of tax non-compliance. Thus, the tax administration effort is placed on devising strategies and structures ensuring that compliance with tax legislation is maximized.

A tax administration can utilize many different measures to increase and maintain taxpayer compliance. These include taxpayer education, written communication, routine visits, audit visits and legal measures; the choice of action depends on the specific characteristics of each taxpayer, including compliance history, as heuristically determined by the experts of the local tax authorities. For instance, a previously compliant taxpayer with no tax liability that did not file the latest tax return may receive a mere reminder letter. In contrast, a taxpayer with a history of non-compliance and significant tax liability should expect a tax audit.

The immensity of the workload that tax administrations are confronted with renders the effective achievement of their mission a rather elusive task if performed completely manually. Tax administrations are under constant pressure from governments to achieve more with less resources. Tax auditors constitute a very scarce resource and must be deployed with caution in order to maximize return. As only a limited number of audits can be

☆ No author associated with this paper has disclosed any potential or pertinent conflicts which may be perceived to have impending conflict with this work. For full disclosure statements refer to https://doi.org/10.1016/j.knosys.2019.105048.

* Corresponding author.
E-mail addresses: cs.kleanthous@edu.cut.ac.cy (C. Kleanthous), sotirios.chatzis@cut.ac.cy (S. Chatzis).

https://doi.org/10.1016/j.knosys.2019.105048
0950-7051/

performed annually, tax administrations usually target cases that are expected to generate the highest audit yield.

Therefore, to facilitate this strenuous audit selection procedure, tax administrations have long relied on automation. Historically, this was driven from rules-based systems, with the rules heuristically created by experts to classify taxpayers risk and select audit cases [3]. Unfortunately, the rules-based approach is a laborious and time-consuming process. Rules are created based on the experience and expectations (bias) of tax experts. In many cases, rules-based approaches may incorporate hundreds of rules for each specific application area. Then, decision depends upon the number of rules that "fire" in each case. A taxpayer gets audited if the number of rules that fire exceeds a heuristically predefined threshold.

From this description, it can be easily observed that rules-based systems heavily rely on subjective, complicated, error prone and certainly incomplete sets of rules and associated "firing" thresholds. Thus, tax administrations have recently started to consider alternative options. In this context, adoption of advanced data analytics and machine learning rises as a promising option [4]. This is especially the case since tax departments have access to a vast amount of data associated with each taxpayer (e.g., filed VAT returns). Appropriately leveraging this resource may open new avenues for automating the audit selection process, with the goal of targeting cases that may offer high audit yields.

The Organization for Economic Co-operation and Development (OECD) has published a guidance note [5] where data mining, commonly defined as the discovery of models for data, is suggested as a key supplement to the work of the tax experts in identifying non-compliant taxpayers. Machine learning models hold great promise towards the achievement of this vision of next-generation automated and reliable audit case selection. Typically, machine learning models are trained under the "supervised" learning paradigm. This requires availability of data stemming from examples of both compliant and non-compliant taxpayers, and can result in high detection accuracy, contingent upon the proper selection of the machine learning model. Unfortunately, though, available audit data are seldom sufficiently large for developing strong supervised models. Even worse, tax administrations do not publish taxpayer data or taxpayer audit results because they are confidential. As a consequence, researchers not affiliated with a tax authority have very limited capacity to build realistic machine learning models for tax audit selection. This considerably limits the innovation potential in the field.

As a solution to the lack of adequate audit cases that can be used as "labeled" training data for supervised learning algorithm development, several researchers have considered the utilization of unsupervised learning algorithms, specifically clustering techniques. This class of algorithms essentially performs data grouping on the basis of some similarity criterion; the premise is that taxpayers that evade similarly large amounts should group together. However, the efficacy of these approaches relies unacceptably heavily on the appropriateness of the employed similarity criterion. As such, from the perspective of tax authorities, these methods suffer from a considerable lack of trustworthiness.

In this paper, we focus on the problem of VAT audit case selection; that is, the problem of accurately detecting in the immense pool of VAT registered taxpayers those that evade the highest tax amounts. Then, by targeting the limited number of audits that a Tax Department can feasibly perform to these taxpayers, we enable the optimization of the confirmed and recovered amounts of evaded VAT (audit yield), in line with the Tax Department objectives. For the first time in the literature and the state practice alike, we offer an innovative machine learning approach that combines the benefits of both the supervised and unsupervised learning paradigms: Our goal is to allow for yielding dependable predictions based on learning from only a limited number of audit cases ("labeled" data), combined with the large corpora of filed VAT returns that we have access to ("unlabeled" data).

To this end, we make use of the most cutting-edge developments in deep learning [6]. The deep learning breakthrough has enabled a leap-forward in the efficacy of machine learning algorithms. This is mainly due to the fact that it obviates the need of coming up with extensive sets of features for representing the available data. On the contrary, conventional feature engineering is replaced with the capacity to infer hierarchies of robust trainable feature vectors, that can yield optimal representation performance in the available training data. Even more importantly, this procedure can be effected using unlabeled training examples. Labeled data are only needed for training a penultimate classification layer of the developed model. This way, by exploiting the inferred feature representations, effective classification performance can be obtained with only limited training data availability. Therefore, it is no coincidence that challenging real-world applications such as image and video understanding, as well as natural language understanding and generation, have reaped significant benefits by utilizing deep learning [7].

Inspired from these advances, in this work we address VAT audit case selection by devising a novel Gated Mixture Variational Autoencoder network. The deep network-driven nature of the devised model allows for it to be presented with *raw* VAT return data and learn to infer representative features without any need of labeled training examples. At the same time, the model can inherently learn to use the inferred features so as to perform case classification (thus, VAT audit case selection), by using the available labeled data; that is, past audited cases and respective outcomes (high/low audit yield). In addition, our use of variational inference algorithms for model training offers a high level of model regularization, thus reducing overfitting tendencies. We developed and deployed our approach in collaboration with the Cyprus Tax Department, which provided the training data and facilitated performance evaluation. Our method completely outperformed both the currently used rules-based systems, as well as popular machine learning alternatives, offering to the Tax Department a groundbreaking solution for VAT audit case selection.

The remainder of this paper is organized as follows: In the following Section, we provide a brief overview of the related work. In Section 3, we introduce our approach, elaborate on its rationale, and devise its training and inference algorithms. In Section 4, we elaborate on the development of the method and its experimental deployment with the Cyprus Tax Department. Finally, in the concluding section, we summarize our contribution and outline open issues for further research.

2. Related work

2.1. Rules-based and data mining systems

Many EU-based tax departments have relied on the SAS Enterprise Miner[1] to build VAT audit selection models, under a Sample /Explore /Model /Assess data mining framework. Cases are sampled randomly or in a biased fashion, and an understanding of the data is obtained through exploring the use of statistical analysis. Feature engineering is extensively used; it includes imputing missing data, using logs of numeric variables, and grouping the data in categories. A characteristic example is the Irish Revenue Office, which employs data mining for selecting non-compliant taxpayers with significant tax yield for audit [8]. They use SAS

[1] https://www.sas.com/en_us/software/enterprise-miner.html.

Please cite this article as: C. Kleanthous and S. Chatzis, Gated Mixture Variational Autoencoders for Value Added Tax audit case selection, Knowledge-Based Systems (2019) 105048, https://doi.org/10.1016/j.knosys.2019.105048.

Enterprise Miner and SAS Enterprise Guide, specifically the tools initially developed for credit scoring banks and customers of insurance companies.[2] Other EU-based tax authorities, including Cyprus Tax Authority, employ in-house developed data mining solutions that use open-source software and internally proposed heuristics (rules/thresholds).

2.2. Machine learning approaches

As already discussed in the Introduction, machine learning promises higher performance compared to rules-based approaches and simplistic data mining, since there is no need of creating detailed rules for each and every task; this also saves time for the data analyst and the VAT expert. Besides, the efficacy of a rules-based approach relies on the experts being capable of capturing the most salient possibilities in terms of tax evasion or avoidance cases; this is virtually impossible in any real-world situation.

On the contrary, machine learning models are popular for their capacity to generalize in unforeseen cases which may share some common underlying patterns with previous examples. This generalization capacity may be extremely strong if the postulated model is configured appropriately, e.g. [10]. In this context, supervised learning (classification) algorithms are among the first used for addressing the audit selection task [4]. A successful predictive model must be able to accurately classify the taxpayers in terms of whether a significant VAT audit yield will arise in case of audit. Techniques such as Support Vector Machines (SVMs) [11] and Decision Trees (CART) [12] are widely used to this end.

However, the generalization capacity of supervised techniques is also contingent upon the availability of rich labeled training datasets, that is an adequate number of VAT audit results (high/low labels). In case the number of VAT audits performed is inadequate, model training becomes unsuccessful. Therefore, relying on supervised learning techniques trained using prior audit data poses the danger of creating models with limited generalization capacity. This problem of generalization performance is widely noted in the existing literature, e.g. [4,13], and has greatly discouraged EU-based tax authorities from performing research and development of supervised learning models for VAT audit selection.

To ameliorate these drawbacks, several researchers have resorted to unsupervised learning techniques, namely clustering algorithms. We elaborate on these techniques in the following Section.

2.3. Unsupervised techniques

The main rationale of unsupervised learning application consists in the assumption that entities which perform significant tax evasion may exhibit similar behavioral patterns; hence, a well-designed unsupervised learning algorithm should be capable of clustering them together. The Australian Tax office has been a pioneer in the application of real-time analytics using unsupervised techniques; taxpayers are classified while completing their tax return. They use the 'Nearest Neighbor' algorithm to detect similar taxpayers which are expected to file comparable amounts in their tax returns. Taxpayers filing significantly different amounts than expected are encouraged in real-time to revise their declared amounts. According to the Australian Tax Office, 7% of taxpayers (230,000) made an upward adjustment after receiving this automated message in 2017; this corresponds to a total additional revenue of AU$95 million, without carrying a single audit [14].

The recent work of [4] is an example of utilizing unsupervised machine learning to detect anomalies in clusters formed on tax return data. Anomaly detection in clustering algorithm outcomes has also been considered in [15], where spectral clustering is used, as well as in [16], where manifold learning is employed. Unsupervised learning combined with expert feedback has also been considered in [17]. Matos et al. in [18] combined association rules and dimensionality reduction via Singular Value Decomposition and Principal Component Analysis. The popular k-means clustering algorithm, and its numerous variants, have also been used for tax audit selection [19].

Despite these advances, a direct consequence of not using label information is the fact that there are no right or wrong answers expected by a clustering algorithm, contrary to supervised learning. Thus, it is not straightforward to discern how the algorithm outcomes can be actually used to drive the audit selection process. Sets of heuristics must be devised, which brings again to the fore the risk of subjectivity and limited generalization capacity. Besides, the presence of outliers, which is rather commonplace in real-world data, may catastrophically mislead the vast majority of clustering algorithms [20], thus further undermining their efficacy and reliability as a tax audit selection tool.

Finally, a group of researchers have recently proposed a different solution to the lack of labeled data [21]; this consists in dataset augmentation via Generative Adversarial Networks [22]. This is an attempt to train a reliable generator of synthetic labeled data to augment the available labeled datasets with. However, the performance improvement obtained in this work was far from satisfactory. This was actually more than expectable: if the available labeled dataset is insufficient for training a classifier (and augmentation is needed), one would expect it to also be insufficient for generator network training.

3. Methodology

3.1. Motivation

This work takes a different route in the effort of addressing the limited labeled training data availability that plagues the application of supervised machine learning models to automated VAT audit selection. Specifically, we pioneer the utilization of the semi-supervised machine learning paradigm, with strong inspiration from recent developments in the field of deep learning [23].

Semi-supervised learning in its simplest form assigns predicted labels to the unlabeled data and incorporates them in the training set [24]. A number of repetitions is performed until a preset convergence criterion is met. However, this procedure may result in poor predictions being reinforced. More advanced procedures that ameliorate this risk employ graph-based methods that create a graph connecting similar observations; when a minimum energy configuration is found, the label information is propagated between labeled and unlabeled nodes [25]. The inherent limitation of this paradigm is limited scalability [23]. Another example is the Transductive Support Vector Machine (TSVM) [26] semi-supervised classification model; this enhances basic SVM's so as to use the minimum number of predicted output labels which are near the margin.

A groundbreaking paradigm that bears great promise towards resolving these issues is deep learning. Specifically, deep networks can be simultaneously trained under both the supervised and unsupervised learning paradigms. For instance, autoencoders [27] are deep network configurations that are typically trained in an unsupervised fashion, on the basis of an observation representation (encoding) and reconstruction error criterion.

[2] Notably, SAS and IBM lead the market share of predictive analytics tools by revenue volume, with a combined market share of 43% [9].

Please cite this article as: C. Kleanthous and S. Chatzis, Gated Mixture Variational Autoencoders for Value Added Tax audit case selection, Knowledge-Based Systems (2019) 105048, https://doi.org/10.1016/j.knosys.2019.105048.

However, they can also be used as an intricate part of a supervised deep classifier, so as to facilitate training of the network intermediate layers by exploiting vast amounts of unlabeled data [28].

Semi-supervised learning has already been successfully applied to fraud detection tasks. For instance, Zhang in [29] proposed a binary classification of tax declarations (fraudulent/not fraudulent) using unlabeled and expert-marked data to fine-tune weights of a deep network. On a different vein, two subsets of credit card transactions were used in [30] to identify suspicious transactions. However, VAT audit selection has never been addressed before.

These facts constitute a major source of inspiration for this work.

3.2. Model formulation

Our work is an attempt to answer the following fundamental question: **"Can tax administrations leverage non-audited filed VAT returns to accurately predict whether a prospective audit will achieve high or low yield?"** To obtain a convincing answer, we develop a tailor-made deep learning model, whereby we cast the problem into classification as cases of high or low potential audit yield. Then, we address the introduced problem by leveraging the latest advances in the field of autoencoder deep networks, namely variational autoencoders (VAEs) [27,31,32].

Initially, we process the raw data included in the taxpayers quarterly VAT returns to obtain the observations presented to the network. The obtained measurements comprise: (i) economic activity type, classified according to the Eurostat NACE classification[3]; (ii) district codes; (iii) type of taxpayer (physical, legal); (iv) *raw* declared amounts, including VAT due/local sales, VAT due/EU purchases, VAT refundable (purchases), VAT payable, net value of sales, net value of purchases, value of zero-rated sales, value of purchases from EU (goods and services), and value of sales to EU (goods and services). These raw measurements used to quantitatively describe our data were selected based on the advice of experienced field auditors, who have devised the heuristic rules currently used by the Cyprus Tax Department.

Eventually, we end up with a total of 47 raw measurements that constitute the observed data fed into the devised model. As labels associated with these observations, we use the corresponding audit outcomes, if an audit has been performed. Apparently, since only a small fraction of the filed returns are audited, most of the available data points are unlabeled.

The so-obtained observations are presented to an *encoder* network; this splits into two parts, with the first being an *intermediate* dense layer that comprises 40 ReLU units. Drawing from the recent advances in the field of variational autoencoders, e.g. [31], this encoder network facilitates the modeling process by learning to infer a high-level representation of the observed measurements. This representation is more useful for the classification process compared to the measurements themselves [31,32]. As shown in Fig. 1, the intermediate layer of the encoder network is followed by a second part comprising two distinct *subencoders* that work in tandem. This is a radically novel modeling selection adopted in our work, which differentiates it from the existing literature. Both these subencoders are presented with the 40-dimensional output of the intermediate layer, and generate a final 20-dimensional (latent) vector, again obtained from ReLU nonlinearities. These 20-dimensional latent vector representations (encodings) are propagated to the subsequent parts of the proposed model.

[3] https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF.

The rationale behind this novel configuration of the encoder of the devised model is motivated by a key observation; the two modeled classes (high/low yield) are expected to entail significantly different patterns of latent underlying dynamics. Hence, it is plausible that each class can be adequately and effectively modeled by means of distinct, and different, encoder distributions. We posit that learning these two distinct distributions may be best facilitated by using two subencoders. The distinct subencoder parts allow for differentiation, while the common anterior encoder part enforces our expectation that the two learned encoding distributions share some correlation.

At this point, we introduce another key modeling principle of our method. We consider that the output units of the subencoders are of a stochastic nature; specifically, we consider stochastic outputs, say $\tilde{\boldsymbol{z}}$ and $\hat{\boldsymbol{z}}$, with Gaussian (posterior) densities. This assumption renders our model a variational autoencoder (as opposed to a conventional autoencoder model). We strategically select to adopt the variational inference framework in developing our autoencoder model, as it is well-understood to allow for significantly improved generalization capacity and reduced overfitting tendencies [31]. Hence, what the postulated subencoders actually compute are the means, $\tilde{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\mu}}$, as well as the (diagonal) covariance matrices, $\tilde{\boldsymbol{\sigma}}^2$ and $\hat{\boldsymbol{\sigma}}^2$, that parameterize these Gaussian posteriors. On this basis, the actual subencoder output vectors, $\tilde{\boldsymbol{z}}$ and $\hat{\boldsymbol{z}}$, are sampled each time from the corresponding (inferred) Gaussian posteriors.

Under this mixture model formulation, we need to establish an effective mechanism for inferring which observations (i.e., analyzed VAT returns) are more likely to match the learned distribution of each component subencoder. In layman terms, this can be considered to be analogous to a (soft) classification mechanism differentiating between audit cases of high and low potential yield. This mechanism can be obtained by computation of the posterior distribution of mixture component membership (also known as "responsibility" in the literature of finite mixture models [33]). This is also needed for effectively selecting between the samples of $\tilde{\boldsymbol{z}}$ or $\hat{\boldsymbol{z}}$, at the output of the encoding stage of the devised model, that will be propagated to the subsequent model components.

To allow for inferring this posterior distribution, in this work we postulate a gating network. This is a dense-layer network, presented with the same 40-dimensional intermediate representation, $h()$, as the two postulated subencoders, and using a sigmoid activation function. It is trained alongside the rest of the model, and it is the only part of the model that requires availability of labeled data for its effective training. Thus, under this model construction, the needs of our approach in labeled data availability are considerably reduced.

To conclude the formulation of the proposed model, we need to postulate an appropriate decoder distribution, and a corresponding network that infers it. In this work, we opt for a simple dense-layer neural network, which is fed with the (sampled) output of the postulated finite mixture model encoder, and attempts to reconstruct the original raw measurements. Specifically, we postulate a network comprising one hidden layer with 40 intermediate ReLU units.

Let us denote as $\boldsymbol{x}_n$ the set of observable measurements pertaining to the nth available VAT return. Then, based on the above description, the encoder distribution of the postulated model reads

$$q(\boldsymbol{z}_n|\boldsymbol{x}_n) = q(\tilde{\boldsymbol{z}}_n|\boldsymbol{x}_n)^{q(c_n=1|\boldsymbol{x}_n)} q(\hat{\boldsymbol{z}}_n|\boldsymbol{x}_n)^{q(c_n=0|\boldsymbol{x}_n)} \quad (1)$$

Here, $\boldsymbol{z}_n$ is the output of the encoding stage of the proposed model that corresponds to $\boldsymbol{x}_n$, $\tilde{\boldsymbol{z}}_n$ is the output of the first subencoder, corresponding to the high yield class, $\hat{\boldsymbol{z}}_n$ is the output of the second subencoder, corresponding to the low yield class, and c_n

Please cite this article as: C. Kleanthous and S. Chatzis, Gated Mixture Variational Autoencoders for Value Added Tax audit case selection, Knowledge-Based Systems (2019) 105048, https://doi.org/10.1016/j.knosys.2019.105048.

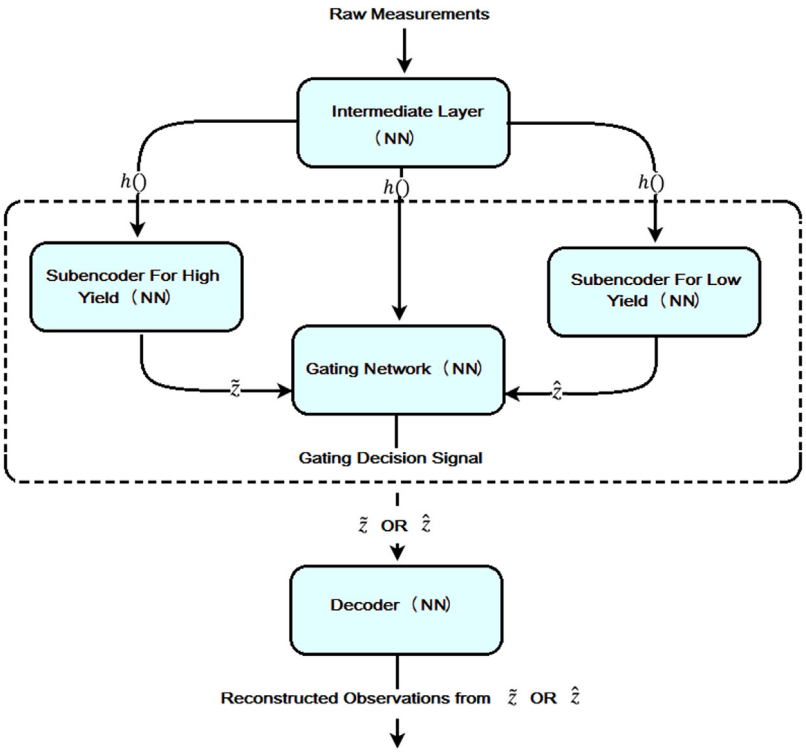

Fig. 1. Overview of the proposed model.

is a latent variable indicator of whether $\boldsymbol{x}_n$ belongs to the high yield class or not. We also postulate

$$q(\tilde{\boldsymbol{z}}_n|\boldsymbol{x}_n) = \mathcal{N}(\tilde{\boldsymbol{z}}_n|\tilde{\boldsymbol{\mu}}(\boldsymbol{x}_n;\tilde{\boldsymbol{\theta}}), \text{diag}\,\tilde{\boldsymbol{\sigma}}^2(\boldsymbol{x}_n;\tilde{\boldsymbol{\theta}})) \tag{2}$$

$$q(\hat{\boldsymbol{z}}_n|\boldsymbol{x}_n) = \mathcal{N}(\hat{\boldsymbol{z}}_n|\hat{\boldsymbol{\mu}}(\boldsymbol{x}_n;\hat{\boldsymbol{\theta}}), \text{diag}\,\hat{\boldsymbol{\sigma}}^2(\boldsymbol{x}_n;\hat{\boldsymbol{\theta}})) \tag{3}$$

Here, the $\tilde{\boldsymbol{\mu}}(\boldsymbol{x}_n;\tilde{\boldsymbol{\theta}})$ and $\tilde{\boldsymbol{\sigma}}^2(\boldsymbol{x}_n;\tilde{\boldsymbol{\theta}})$ are outputs of the deep neural network that corresponds to the high yield class subencoder, with parameters set $\tilde{\boldsymbol{\theta}}$. Similarly, the $\hat{\boldsymbol{\mu}}(\boldsymbol{x}_n;\hat{\boldsymbol{\theta}})$ and $\hat{\boldsymbol{\sigma}}^2(\boldsymbol{x}_n;\hat{\boldsymbol{\theta}})$ are outputs of the deep neural network that corresponds to the low yield class subencoder, with parameters set $\hat{\boldsymbol{\theta}}$.

The posterior distribution of mixture component allocation, $q(c_n|\boldsymbol{x}_n)$, which is parameterized by the aforementioned gating network, is a simple Bernoulli distribution that reads

$$q(c_n|\boldsymbol{x}_n) = \text{Bernoulli}(\varpi(h(\boldsymbol{x}_n);\boldsymbol{\varphi})) \tag{4}$$

Here, $\varpi(h(\boldsymbol{x}_n);\boldsymbol{\varphi}) \in [0, 1]$ is the output of the gating network, with trainable parameters set $\boldsymbol{\varphi}$. This infers the probability of $\boldsymbol{x}_n$ belonging to the high yield class.

Lastly, the postulated decoder distribution reads

$$p(\boldsymbol{x}_n|\boldsymbol{z}_n) = \mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}(\boldsymbol{z}_n;\boldsymbol{\phi}), \text{diag}\,\boldsymbol{\sigma}^2(\boldsymbol{z}_n;\boldsymbol{\phi})) \tag{5}$$

where the means and diagonal covariances, $\boldsymbol{\mu}(\boldsymbol{z}_n;\boldsymbol{\phi})$ and $\boldsymbol{\sigma}^2(\boldsymbol{z}_n;\boldsymbol{\phi})$, are outputs of a deep network with trainable parameters set $\boldsymbol{\phi}$, configured as described previously.

3.3. Model training

Let us consider a training dataset $X = \{\boldsymbol{x}_n\}_{n=1}^N$ that consists of N filed VAT returns. A small subset, X^l, of size M of these samples is considered to be labeled, with corresponding labels set $Y = \{y_m\}_{m=1}^M$. That is, these VAT returns triggered an audit, which may have generated a high or low audit yield ($y_m = 1$ and $y_m = 0$, respectively). Then, following the VAE literature [27], model training is performed by maximizing the evidence lower bound (ELBO) of the model over the parameters set $\{\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}, \boldsymbol{\varphi}, \boldsymbol{\phi}\}$. The ELBO of our model reads:

$$\log p(X) \geq \mathcal{L}(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}, \boldsymbol{\varphi}, \boldsymbol{\phi}|X) = -\sum_{n=1}^{N} \text{KL}\big[q(\boldsymbol{z}_n|\boldsymbol{x}_n) \parallel p(\boldsymbol{z}_n)\big] + \gamma \sum_{n=1}^{N} \mathbb{E}[\log p(\boldsymbol{x}_n|\boldsymbol{z}_n)] + \sum_{\boldsymbol{x}_m \in X^l} \log q(c_m = y_m|\boldsymbol{x}_m) \tag{6}$$

Here, $\text{KL}\big[q \parallel p\big]$ is the KL divergence between the distribution $q(\cdot)$ and the distribution $p(\cdot)$, while $\mathbb{E}[\cdot]$ is the (posterior) expectation of a function w.r.t. its entailed random (latent) variables. Note also that, in the ELBO expression (6), the introduced hyperparameter γ is a simple regularization constant, employed to ameliorate the overfitting tendency of the postulated decoder networks, $p(\boldsymbol{x}_n|\boldsymbol{z}_n)$. We have noticed that this simple trick yields a significant improvement in generalization capacity.

In Eq. (6), the posterior expectation of the log-likelihood term $p(\boldsymbol{x}_n|\boldsymbol{z}_n)$ cannot be computed analytically, due to the nonlinear form of the decoder. Hence, we must approximate it by drawing Monte-Carlo (MC) samples from the posterior (encoder) distributions (2)–(3). However, MC gradients are well-known to suffer from high variance. To resolve this issue, we utilize a smart re-parameterization of the drawn MC samples. Specifically, following the related derivations in [27], we express these samples in the form of a differentiable transformation of an (auxiliary) random noise variable $\boldsymbol{\epsilon}$; this random variable is the one we actually draw MC samples from:

$$\tilde{\boldsymbol{z}}_n^{(s)} = \tilde{\boldsymbol{\mu}}_n + \tilde{\boldsymbol{\sigma}}_n \cdot \boldsymbol{\epsilon}_n^{(s)}, \;\; \boldsymbol{\epsilon}_n^{(s)} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}) \tag{7}$$

$$\hat{\boldsymbol{z}}_n^{(s)} = \hat{\boldsymbol{\mu}}_n + \hat{\boldsymbol{\sigma}}_n \cdot \boldsymbol{\epsilon}_n^{(s)} \tag{8}$$

Hence, such a re-parameterization reduces the computed expectations into averages over samples from a random variable with

Please cite this article as: C. Kleanthous and S. Chatzis, Gated Mixture Variational Autoencoders for Value Added Tax audit case selection, Knowledge-Based Systems (2019) 105048, https://doi.org/10.1016/j.knosys.2019.105048.

low (unitary) variance, ϵ. This way, by maximizing the obtained ELBO expression, we yield low-variance estimators of the sought (trainable) parameters, under some mild conditions [27]. We perform the maximization process of $\mathcal{L}(\tilde{\boldsymbol{\theta}}, \hat{\boldsymbol{\theta}}, \boldsymbol{\varphi}, \boldsymbol{\phi}|X)$ by resorting to AdaGrad [34].

3.4. Prediction generation

To predict the class (high/low yield) of a VAT audit case (filed return data), $\boldsymbol{x}_n$, we compute the mixture assignment posterior distribution $q(c_n|\boldsymbol{x}_n)$, inferred via the postulated gating network, $\varpi(h(\boldsymbol{x}_n); \boldsymbol{\varphi})$. On this basis, assignment is performed to the high-yield class if $\varpi(h(\boldsymbol{x}_n); \boldsymbol{\varphi}) > 0.5$.

4. Method deployment

4.1. Development process

The motivating force of this work has been the pressing need to reliably automate the VAT audit selection process for the Cyprus Tax Department. As such, development of the devised Gated Mixture Variational Autoencoder was performed with their close collaboration. Specifically, we gathered over 1,000,000 filed VAT returns as unlabeled data and over 10,000 audited VAT returns as labeled data.[4] These constitute nearly all the VAT returns of the last six years. Following the instructions of the Tax Department, and to best facilitate their needs, we have considered three *alternative model configurations*: (i) learning to detect potential audit yields exceeding €100; (ii) exceeding €75; (iii) exceeding €67; and (iv) exceeding €50.

We used this dataset to both train and evaluate our model and the considered competitors. Specifically, training was performed using the whole set of unlabeled data, and a fraction of the labeled ones under a 4-fold stratified cross-validation rationale; the rest of the available labeled data was used for model evaluation (in each iteration of the 4-fold cross-validation process).

The proposed approach was implemented in Python, using the TensorFlow library [35]. The developed models were run on a Desktop PC hosting an off-the-shelf NVIDIA 10 series Graphic Processing Unit. To perform model training, we used $S = 10$ drawn MC samples, $\boldsymbol{\epsilon}^{(s)}$; we found that increasing this value does not yield any statistically significant accuracy improvement, despite the associated increase in computational costs.

To enable automatic determination of the optimal selection of model hyperparameters, which in the case of deep networks includes the number of hidden layers, the number of units in each layer, the employed nonlinearities, the used batch-size, and the selection of the Dropout and learning rates, we resorted to Neural Architecture Search (NAS) [36] which is now the state-of-the-art paradigm in Machine Learning for hyperparameter selection. Model training was performed via Adagrad.

4.2. The disappointment of supervised learning: Evaluation of a simple dense network alternative

Initially, we examined the efficacy of a state-of-the-art alternative to our approach. Specifically, we considered a conventional deep network which constitutes a supervised learning *alternative* to our approach. We used the available labeled data points to train this deep learning alternative, and resorted to NAS to determine its optimal configuration; this yielded two dense hidden layers with 40 and 20 ReLU units, respectively, regularized via Dropout [37] with rate 0.2.

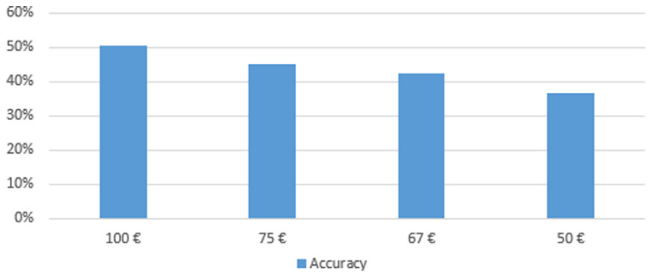

Fig. 2. Supervised model: Obtained accuracy for the audit yield outcomes most typically considered by the Cyprus Tax Authority.

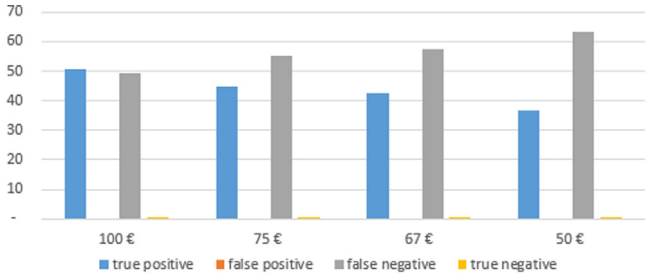

Fig. 3. Supervised model: Confusion matrices for the audit yield outcomes most typically considered by the Cyprus Tax Authority.

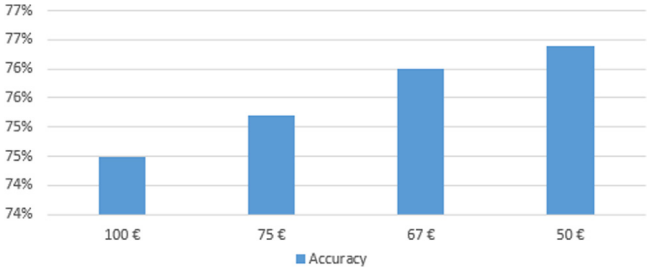

Fig. 4. Proposed system: Obtained accuracy for the audit yield outcomes most typically considered by the Cyprus Tax Authority.

As we illustrate below, the obtained results were far from encouraging; specifically, they were close to the random performance model accuracy (Fig. 2) across all the four tested model configurations (€100, 75, 67 and 50). The confusion matrices (Fig. 3) across all model configurations were also disappointing, as the outcomes are clearly imbalanced. This proves that, with this limited availability of labeled samples, a state-of-the-art supervised model fails to learn any meaningful classification pattern.

4.3. The promise of semi-supervised deep learning models

Subsequently, we proceeded to implement and deploy our proposed Gated Mixture Variational Autoencoder, using the full available dataset (both labeled and unlabeled data points). To obtain a statistically significant evaluation outcome, we performed 4-fold stratified cross-validation, as previously. In addition, to obtain some *comparative results*, we also developed and deployed an existing state-of-the-art *competitor*, namely the M1+M2 semi-supervised deep learning model introduced in [38]. This model comprises a variational autoencoder with dense-network encoder and decoder, combined with a softmax classification layer; it has

[4] Note that, since actual VAT returns and VAT audit results from the Cyprus Tax Department are used, we are restricted from disclosure of the used data and codes, as they constitute privileged information.

Please cite this article as: C. Kleanthous and S. Chatzis, Gated Mixture Variational Autoencoders for Value Added Tax audit case selection, Knowledge-Based Systems (2019) 105048, https://doi.org/10.1016/j.knosys.2019.105048.

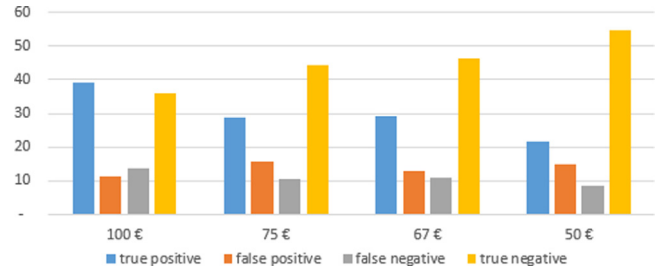

Fig. 5. Proposed system: Confusion matrices for the audit yield outcomes most typically considered by the Cyprus Tax Authority.

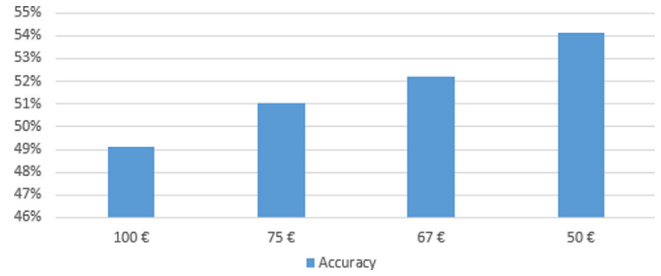

Fig. 6. M1+M2 model: Obtained accuracy for the audit yield outcomes most typically considered by the Cyprus Tax Authority.

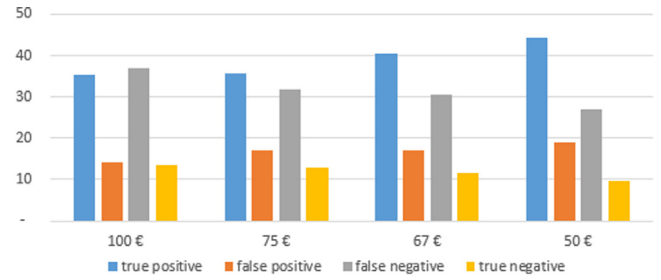

Fig. 7. M1+M2 model: Confusion matrices for the audit yield outcomes most typically considered by the Cyprus Tax Authority.

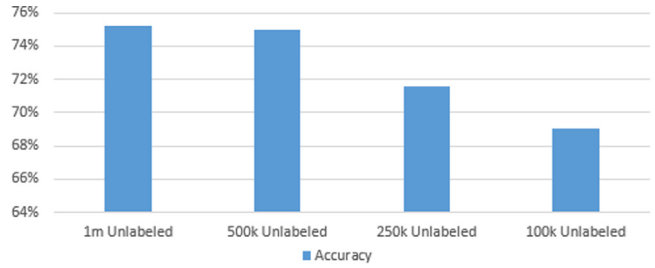

Fig. 8. Proposed system: Accuracy variation by altering the number of used unlabeled data points (€100 audit yield detection).

Fig. 9. Proposed system: Confusion matrix variation by altering the number of used unlabeled data points (€100 audit yield detection).

been shown to greatly and consistently outperform all popular semi-supervised classification alternatives, including the popular TSVM [26]. NAS yielded a M1+M2 configuration comprising 40 intermediate units and 20-dimensional latent vectors; exactly the same configuration NAS obtained for our approach.

Fig. 4 depicts the detection accuracy obtained by our proposed system for the audit yield outcomes most typically considered by the Cyprus Tax Authority; Fig. 5 shows the corresponding confusion matrices. As we observe, despite the limited availability of labeled samples, our approach yields quite a high accuracy level across all the considered scenarios. This represents a dramatic improvement over the supervised learning alternative, providing strong evidence of the efficacy of our proposed approach and the importance of appropriately leveraging unlabeled data in the context of our addressed problem.

Further, we provide the corresponding evaluation outcomes pertaining to the considered M1+M2-based alternative. These are shown in Figs. 6 and 7, respectively. It becomes apparent that the M1+M2 algorithm is incapable of yielding any meaningful performance outcome, as it has barely managed to exceed 50% in all scenarios. This provides indisputable evidence of the superiority of our modeling approach, including both the proposed split of the encoder module, as well as the use of the gating network (classifier) as an integral part of the variational autoencoder. Therefore, we deduce that resorting to a state-of-the-art semi-supervised learning algorithm does not guarantee effective exploitation of unlabeled data. Addressing the task at hand requires significant expertise and understanding of the problem, combined with the capacity to build upon and extend the state-of-the-art in machine learning.

4.4. Ablation study

Finally, to obtain a deeper understanding of how unlabeled training data availability facilitates the modeling performance of the proposed Gated Mixture Variational Autoencoder model, we performed an extensive ablation study. Focusing on the target audit yield of €100 outcome, we repeated our evaluation by reducing the number of used unlabeled training data points. Specifically, we examined three different test cases, where we used a randomly sampled fraction of the unlabeled data points comprising 500k, 250k and 100k samples, respectively.

The obtained results are provided in Fig. 8 (accuracy) and Fig. 9 (confusion matrices). We observe that performance remains robust as we decrease the number of unlabeled data points by 50% (500k unlabeled data points), but deteriorates if we reduce these even further. Characteristically, when only a 10% of the originally available unlabeled data is used, the accuracy drops by 7 percentage points. However, it remains profoundly better than the M1+M2 model and the evaluated supervised alternative. This constitutes conspicuous empirical evidence of the solid methodological foundation and versatility of the devised solution to the addressed problem of VAT audit case selection.

4.5. System adoption

The previous results strongly support the efficacy of the proposed system. As the €100 baseline is the targeted audit yield threshold for the rule-based systems currently used by the Cyprus Tax Department, it is important to stress that the obtained performance outcome offers an unprecedented level of reliability for tax auditors. Figs. 10 and 11 summarize how strong the improvement of our approach is over supervised techniques. In

Please cite this article as: C. Kleanthous and S. Chatzis, Gated Mixture Variational Autoencoders for Value Added Tax audit case selection, Knowledge-Based Systems (2019) 105048, https://doi.org/10.1016/j.knosys.2019.105048.

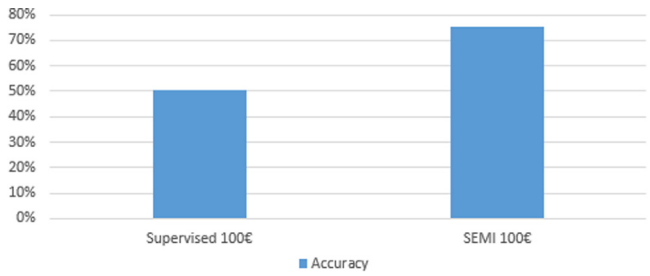

Fig. 10. Accuracy: Supervised vs. Semi-supervised model.

Fig. 11. Confusion matrices: Supervised vs. Semi-supervised model.

addition, we emphasize that the *currently used rules-based systems* developed by the Cyprus Tax Department for assisting the VAT audit case selection process achieve a success rate that *fluctuates between 60–65%* (depending on seasonality effects). Note also that prediction generation using our model requires only feedforward computation encompassing the anterior part of the encoder and the gating network; as such, predictions are obtained momentarily. Hence, our study represents a giant leap-forward towards the goal of more effective and targeted VAT audit selection. Its full deployment, which remains open to further (longer-term) performance confirmation, is expected to eventually catalyze a significant reduction in Cyprus VAT-gap.

5. Conclusions and future work

The mission of the Cyprus Tax Department is the *"consistent application of the laws, ensuring fair taxation in a way that enhances the confidence of the taxpayer, the minimization of tax evasion and the effective collection of tax revenues of the state with the least possible cost"*. [39]. As VAT is one of the major sources of tax evasion [2], this work attempted to provide an effective automated solution to the problem of VAT audit case selection. This was expected to greatly facilitate the Tax Department in its effort to reduce tax evasion, by utilizing its limited resources (experienced auditors) to target cases with high audit yield.

To address this challenge, we devised a novel Gated Mixture Variational Autoencoder model, drawing from the latest advances in deep learning and model regularization via variational inference. Our goal was to develop a full methodological pipeline that obviates the need for tax experts to create hundreds of detailed rules; a procedure extremely time-consuming, costly, and disturbingly imprecise. At the same time, our approach was designed to make the most out of the available audit data, taking under consideration that their availability is too limited for a supervised learning algorithm to achieve satisfactory performance. Our devised innovative model combined a supervised component, namely a gating network discerning between alternative decisions, and an unsupervised one, namely a variational autoencoder. The gating network generates predictions for each case (high/low yield), while both components are trained in tandem.

At this point, we emphasize that in this work we cast the VAT audit case selection process as a binary classification problem. Our treatment was motivated from the actual current business practice of the Cyprus Tax Department. Indeed, attempting to frame VAT audit case selection into a more general multi-label paradigm posed the risk of being too disruptive to the current practice; this could hinder the eventual uptake of our innovative solution. Therefore, considering such a multi-label modeling scheme is beyond the scope of this paper. However, we stress that utilization of our model to address multi-label classification problems is straightforward; one needs to split the encoder module to as many subencoders as the number of considered classes, and replace the employed sigmoid gating network with a softmax that now parameterizes a Discrete posterior, as opposed to the Bernoulli in (4). We admit that such a network design is not scalable to many classes. However, Tax Departments are interested in predicting the additional revenue generated by a prospective audit, rounded to the nearest tens of thousands of euros; therefore, in the worst-case scenario, the number of classes one needs to model is limited to just a few.

We developed and deployed an experimental prototype of our system by making use of more than one million quarterly VAT returns filed in the last years, as well as 10,000 associated audit outcomes. Eventually, not only did our approach significantly surpass the high-yield audit case selection success rate of the currently used rules-based systems, which stands at 60%–65% of the audited cases; even more profoundly, it completely outperforms popular alternative machine learning algorithms, including state-of-the-art deep networks and Transductive SVM's, which virtually failed to obtain any meaningful outcome.

These conspicuous outcomes have prodded the Tax Department to perform a set of follow-up evaluation cycles for performance verification purposes. The ultimate vision is to fully integrate the system into the Department's standard VAT audit selection practices, replacing the rules-based systems currently used. In addition, we examine how this system can be leveraged to address other sources of tax evasion. Indeed, it is common practice for tax administrations to cross-validate and reconcile items declared in the tax returns of corporation tax and VAT, like revenue; a taxpayer who filed substantially different revenue amounts should expect an inquiry from the tax authorities. Therefore, taxpayers who under-declare revenue in their VAT returns are also expected to under-declare revenue for direct taxation purposes, and vice versa, so as to avoid attracting the scrutiny of tax authorities. Since VAT evasion and direct tax evasion are correlated, a model that combines raw data from both VAT returns and direct tax returns and performs joint audit case selection for both should yield higher accuracy compared to models addressing VAT and direct taxes separately. This remains to be confirmed in the context of our future research endeavors.

Finally, it is worth to note that the greatest achievement of this project was the stimulation of interest within the Cyprus Tax Department for developing in-house state-of-the-art deep learning tools. Indeed, the success of our project has fostered a pro-research culture, which is especially favorable to further investment in machine learning, in close collaboration with the Academia.

Acknowledgments

This work was supported by the Cyprus Tax Department. We thank the Commissioner of the Department, Mr. Yiannis Tsangaris, for his invaluable support and investment in the pursued state-of-the-art technology.

Please cite this article as: C. Kleanthous and S. Chatzis, Gated Mixture Variational Autoencoders for Value Added Tax audit case selection, Knowledge-Based Systems (2019) 105048, https://doi.org/10.1016/j.knosys.2019.105048.

References

[1] European Commission Directorate-General for Taxation and Customs Union, Taxation trends, 2018.
[2] European Commission Directorate-General for Taxation and Customs Union, Study and reports on the vat-gap in the eu-28 member states:2018 finalreport, TAXUD/2015/CC/131 (2018).
[3] R.-S. Wu, C.S. Ou, H. ying Lin, S.-I. Chang, D.C. Yen, Using data mining technique to enhance tax evasion detection performance, Expert Syst. Appl. 39 (10) (2012) 8769–8777.
[4] D. de Roux, B. Pérez, A. Moreno, M. del Pilar Villamil, C. Figueroa, Tax fraud detection for under-reporting declarations using an unsupervised machine learning approach, in: Proc. ACM KDD, Vol. 8, ACM, New York, NY, USA, 2018, pp. 215–222.
[5] Forum on Tax Administration / Compliance Sub-group, Compliance Risk Management:Managing and Improving Tax Compliance, Technical Report, ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVEL- OPMENT, 2004.
[6] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, Nature 521 (2015) 436–444.
[7] I. Sutskever, O. Vinyals, Q. V. Le, Sequence to sequence learning with neural networks, in: Proc. NIPS, Vol. 4, 2014.
[8] D. Cleary, Predictive analytics in the public sector: Using data mining to assist better target selection for audit, Electron. J. e-Gov. 9 (2) (2011) 132–140.
[9] D. Vesset, C. Gopal, C.W. Olofson, S. Bond, M. Fleming, D. Schubmehl, Worldwide big data and analytics software 2017 market shares: Healthy growth across the board, IDC's Worldwide Big Data and Analytics Software Taxonomy US42353216 (2017).
[10] L. da Silva, H. Rigitano, R. Carvalho, et al., Bayesian networks on income tax audit selection-a case study of brazilian tax administration, in: Proc. UAI, 2019.
[11] P.-F. Pai, M.-F. Hsu, m.-c. Wang, A support vector machine-based model for detecting top management fraud, Knowl.-Based Syst. 24 (2011) 314–321, http://dx.doi.org/10.1016/j.knosys.2010.10.003.
[12] C.-C. Lin, A. Chiu, S.Y. Huang, D. C. Yen, Detecting the financial statement fraud: The analysis of the differences between data mining techniques and experts' judgments, Knowl.-Based Syst. 89 (2015) http://dx.doi.org/10.1016/j.knosys.2015.08.011.
[13] P. Hájek, R. Henriques, Mining corporate annual reports for intelligent detection of financial statement fraud – a comparative study of machine learning methods, Knowl.-Based Syst. 128 (2017) http://dx.doi.org/10.1016/j.knosys.2017.05.001.
[14] Australian Tax Office, How we use data and analytics, 2018.
[15] K. Nian, H. Zhang, A. Tayal, T. Coleman, Y. Li, Auto insurance fraud detection using unsupervised spectral ranking for anomaly, J. Financ. Data Sci. 2 (2016) 58–75.
[16] C. Olson, K. Judd, J. Nichols, Manifold learning techniques for unsupervised anomaly detection, Expert Syst. Appl. 91 (2017) 374–385.
[17] I. Kose, M. Gokturk, K. Kilic, An interactive machine-learning-based electronic fraud and abuse detection system in healthcare insurance, Appl. Soft Comput. 36 (2015) 283–299.
[18] T. Matos, J. Macedo, J. Monteiro, An empirical method for discovering tax fraudsters, in: The 19th International Database Engineering & Applications Symposium on - IDEAS '15, ACM Press, 2014.
[19] E. Neves, A. Dias, C. Pinto, J. Batista, Signaling tax evasion by using financial ratios and cluster analysis, in: Proc. I2FC 2014, 2014.
[20] S. Chatzis, D. Kosmopoulos, T. Varvarigou, Signal modeling and classification using a robust latent space model based on t distributions, IEEE Trans. Signal Process. 56 (3) (2008).
[21] U. Fiore, A. De Santis, F. Perla, P. Zanetti, F. Palmieri, Using generative adversarial networks for improving classification effectiveness in credit card fraud detection, Inform. Sci. (2017).
[22] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Advances in Neural Information Processing Systems 27, 2014, pp. 2672-2680.
[23] Y.W. Rob Fergus, A. Torralba, Semi-supervised learning in gigantic image collections., 522-530, 2009.
[24] C. Rosenberg, M. Hebert, H. Schneiderman, Semi-supervised self-training of object detection models, in: Proc. IEEE Workshops on Application of Computer Vision, 2005, pp. 29-36.
[25] X. Zhu, Z. Ghahramani, J. D. Lafferty, Semi-supervised learning using gaussian fields and harmonic functions., in: Proc. ICML, 2003, pp. 912-919.
[26] T. Joachims, Transductive inference for text classification using support vector machines, in: Proc. ICML, 2001.
[27] D. Kingma, M. Welling, Auto-Encoding Variational Bayes, in: Proc. ICLR, 2014.
[28] J. Weston, F. Ratle, R. Collobert, Deep learning via semi-supervised embedding, Proc. ICML (2008).
[29] K. Zhang, A. Li, B. Song, Fraud detection in tax declaration using ensemble isgnn, in: Computer Science and Information Engineering, World Congress on, 2009.
[30] V. Zaslavsky, A. Strizhak, Credit card fraud detection using self-organizing maps, Inf. Secur. Int. J. 18 (2006) 48–63.
[31] H. Partaourides, S.P. Chatzis, Asymmetric deep generative models, Neurocomputing 241 (2017) 90–96.
[32] L. Maaløe, C.K. Sønderby, S.K. Sønderby, O. Winther, Auxiliary Deep Generative Models, in: ICML, 2016.
[33] G. McLachlan, D. Peel, Finite Mixture Models, 2000.
[34] J. Duchi, E. Hazan, Y. Singer, Adaptive Subgradient Methods for Online Learning and Stochastic Optimization, JMLR, 2010.
[35] M. Abadi, et al., TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL: http://tensorflow.org/, software available from tensorflow.org.
[36] T. Elsken, J.H. Metzen, F. Hutter, Neural architecture search: A survey, J. Mach. Learn. Res. 20 (2019) 1–21.
[37] N. Srivastava, G.E. Hinton, A. Krizhevsky, I. Sutskever, R.R. Salakhutdinov, Dropout: A Simple Way to Prevent Neural Networks from Overfitting, JMLR, 2014.
[38] D.P. Kingma, S. Mohamed, D. Jimenez Rezende, M. Welling, Semisupervised learning with deep generative models, in: Proc. NIPS, 2014.
[39] Cyprus Tax Department, Vision / mission, 2019. URL: http://www.mof.gov.cy/mof/tax/taxdep.nsf/page04_en/page04_en?opendocument.

Please cite this article as: C. Kleanthous and S. Chatzis, Gated Mixture Variational Autoencoders for Value Added Tax audit case selection, Knowledge-Based Systems (2019) 105048, https://doi.org/10.1016/j.knosys.2019.105048.