Anomaly detection based on interval-valued fuzzy sets: Application to rare sound event detection

Stefano Rovetta[1], Zied Mnasri[1,2], Francesco Masulli[1] and Alberto Cabri[1]

[1]DIBRIS, Università degli studi di Genova, Italy
[2]ENIT, University Tunis El Manar, Tunisia

Abstract

Audio signal processing is moving towards detecting and/or defining rare/anomalous sounds. The application of such an anomaly detection problem can be easily extended to audio surveillance systems. Thus, a rare sound event detection method for road traffic monitoring is proposed in this paper, including detection of hazardous events, i.e., road accidents. The method is based on combining anomaly detection techniques, such as variational autoencoders (VAE) and Interval-valued fuzzy sets. The VAE is used to calculate the reconstruction error of the input audio segment. Based on this reconstruction error, a fuzzy membership function, composed of an optimistic/upper component and a pessimistic/lower component, is calculated. Finally, a probabilistic method for interval comparison is used to calculate the membership score, hence to evaluate the interval-valued fuzzy sets. Finally, classification into anomalous/normal events is obtained by defuzzification. Results show that with a careful parameter setting, the proposed method outperforms the state-of-the-art one-class SVM for anomaly detection.

Keywords

Anomalous sound event detection, anomaly detection, variational autoencoder, fuzzy membership, interval-valued fuzzy sets.

1. Introduction

Anomaly/outlierness/novelty can be defined in different ways [1]: (a) by scarcity, as events occurring with low frequency; (b) by characteristics, as events differing from normal events; (c) by meaning, as events carrying a different meaning than normal events. In the specific application of road audio surveillance, *Anomalous* events are mainly car accidents and other events indicating potential hazards like tire skidding, harsh braking, etc., whereas the *Normal* class covers all other events that may happen on the road, e.g. sound of cars, pedestrians, horn blowing and any other non-hazardous event. This is a particular instance, focused only on anomalous sound categories, of the *sound event detection* (SED) problem.

This problem can be formalized either as a classification task for all perceived events, or as detection of only anomalous/outlier/novel events. In either case, two major issues make this task difficult: First, background noise that fully or partly masks all events, making the resulting signals highly variable; secondly, the rareness of the "interesting" events, such as car accidents, which makes them more difficult to model accurately for scarcity of data.

WILF'21: The 13th international workshop on fuzzy logic and applications, Dec. 20–22, 2021, Vietri sul Mare, Italy

stefano.rovetta@unige.it (S. Rovetta); zied.mnasri@enit.utm.tn (Z. Mnasri); francesco.masulli@unige.it (F. Masulli); alberto.cabri@dibris.unige.it (A. Cabri)

CEUR Workshop Proceedings (CEUR-WS.org)

This implies that not only classes are fuzzy, but the membership itself to any class is affected by a degree of uncertainty. In this case, interval-valued fuzzy sets [2] provide an alternative to crisp clustering or type-1 fuzzy sets, for which uncertainty would have to be precisely modelled, either by identification or, more typically, by arbitrary design.

We state the problem as a classification task based on generative models where the final decision is taken by comparing the inferred interval-valued memberships to the different classes, using a classical metric of interval comparison, named degree of preference [3]. This process allows making the final *Normal*/*Anomalous* class decision without discarding the information about uncertainty expressed by the 2-component fuzzy membership.

2. Related work

Sound event detection (SED) is a relatively young discipline, that has emerged since nearly a decade. Sound recognition methods in general proceed by segmenting signals into fixed-length, possibly overlapping *frames* of relatively short duration (fractions of a second). For anomalous SED, anomaly detection and supervised/unsupervised recognition methods are then applied on the obtained, fixed-size feature vectors.

Several methods have been built around generative models, such as hidden Markov models using Gaussian mixture models. Examples of this approach are Ntalampiras et al. [4] and Heittola et al. [5]. Discriminative methods have also been employed, mainly based on support vector machines (SVM) and neural networks (NN). Examples are Foggia et al. [6] using one-class SVM models for each class. The present authors proposed an ensemble one-class SVM-NN model [7], where one-class SVM detects anomalous data and a NN classifies events.

Unsupervised learning has often been preferred to cope with the issues described. Self-supervised neural networks, such as autoencoders, are well suited to this task. We can mention Wei et al. [8] using a reconstruction autoencoder to compute the anomaly score through metric learning, and Purohit et al. [9] employing a deep autoencoder. Variational autoencoders (VAE) [10], learning a hidden generative representation of the data, are especially interesting.

3. Proposed method

As mentioned, the method uses multiple generative models that learn individual classes, and compares interval-valued memberships by using the degree of preference. It proceeds as follows:

- In the training phase, a dedicated VAE model is learnt on each subset containing only one type of events, i.e. *Normal* or *Anomalous.*
- In the test phase, the RMSE error is calculated between the input, i.e. the feature vector representing the signal, and the reconstructed output of each VAE model.
- For each input signal i, the output error $\epsilon_{i,j}$ of each VAE ($1 \leq i \leq N$ and $1 \leq j \leq C$, for N samples and C classes) is used to compute a fuzzy membership function, that provides a measure of closeness of the signal to the event class on which the VAE model had been trained. In our case, for each input sample $C = 2$ interval membership functions are computed, corresponding to the *Normal* category and the *Anomalous* one.

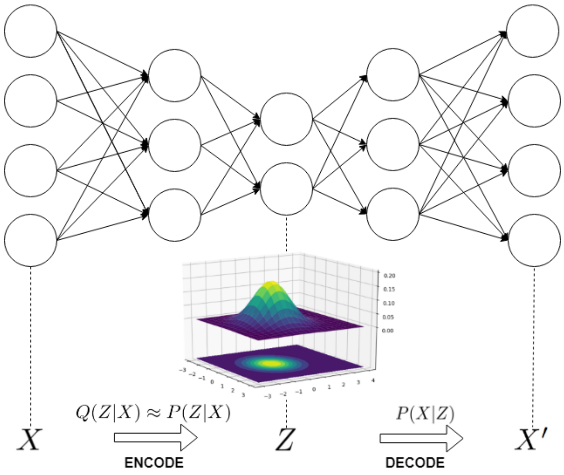

Figure 1: Variational autoencoder

- The membership function associated to each event category, i.e. *Normal*/*Anomalous*, is composed of a low/pessimistic component and an upper/optimistic component, respectively. The values of both components form the interval-valued fuzzy membership function interval (cf. Figure 2).
- Finally, interval comparison is applied using a probabilistic method [11], first to measure the degree of preference of each interval-valued membership function, and subsequently to detect the corresponding event category.

3.1. Variational autoencoder

The variational autoencoder (VAE) is a reconstruction network learning a compressed representation of the input to reconstruct the output. The encoding layer stores the parameters of a probability distribution, e.g., mean and variance, representing the input in a latent space. Then, the decoder uses the probability distribution to generate an approximated reconstruction of the input data. Hence the encoder approximates the probability distribution of the identity function. Given a feature vector X, the VAE aims to find the probability of X with respect to its representation Z,

$$P(X) = \int P(X|Z)P(Z)dZ\,. \tag{1}$$

The network has parameters of $P(Z)$ (average and variance) as its hidden parameters. Using variational inference on a maximum likelihood ojective, the encoder output is trained so that its probability approximates $P(Z|X)$. The reconstruction RMSE can then be obtained as follows:

$$\epsilon = \sqrt{\frac{\sum_{k=1}^{m}(x_k - x'_k)^2}{m}}, \tag{2}$$

where x_i and x'_i $(i = 1, \ldots, N)$ are the input and the output feature vectors for each autoencoder. To compensate for class imbalance, a priori class probabilities are used to compute thresholds.

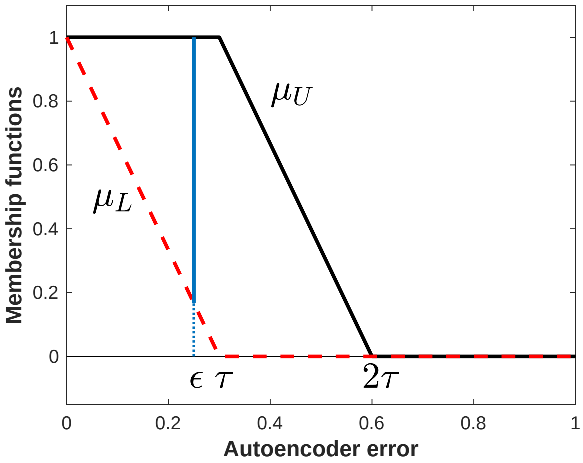

Figure 2: Example of the proposed reconstruction-error-based membership function. Continuous line (μ_U): optimistic membership. Dashed line (μ_L): pessimistic membership. Vertical line at ϵ: interval values of membership corresponding to the reconstruction error ϵ.

In the present work, the VAE employs convolutional layers. The input features are extracted from the spectrogram, i.e. Mel-frequency cepstral coefficients (MFCC) and log-Energy, with their first and second derivatives (Δ and Δ-Δ). The choice of these features is motivated by their proved performance in the state-of-the-art methods of sound event detection [4], in particular road traffic surveillance [6].

3.2. Fuzzy membership function

The membership of each input signal x_i to each event j is computed from on the corresponding VAE's output error $\epsilon_{i,j}$, and its value is the interval between two membership components: a) Pessimistic/Lower membership $\mu_{L,j}$, minimum when the sample is an outlier w.r.t. class j, i.e. $\epsilon_{i,j} > \tau_j$, and b) Optimistic/Upper membership $\mu_{U,j}$, maximum when the sample is classified in class j, i.e. $\epsilon_{i,j} < \tau_j$ (cf. (3)).

$$\mu_{L,j}(\epsilon_{i,j}) = \begin{cases} 1 - \frac{\epsilon_{i,j}}{\tau_j} & \text{if} \quad \epsilon_{i,j} \leq \tau_j \\ 0 & \text{if} \quad \epsilon_{i,j} > \tau_j \end{cases} \qquad \mu_{U,j}(\epsilon_{i,j}) = \begin{cases} 1 & \text{if} \quad \epsilon_{i,j} \leq \tau_j \\ 2 - \frac{\epsilon_{i,j}}{\tau_j} & \text{if} \quad \tau_j < \epsilon_{i,j} \leq 2\tau_j \\ 0 & \text{if} \quad \epsilon_{i,j} > 2\tau_j \end{cases} \tag{3}$$

3.3. Interval comparison

For each class model j, the reconstruction error $\epsilon_{i,j}$ is used to generate the interval membership $M_{i,j} = [\mu_{L,j}(\epsilon_{i,j}), \mu_{U,j}(\epsilon_{i,j})]$. To make the final decision, intervals must be compared for each $j \in \{1, 2\}$. Interval comparison is a particular case of fuzzy number comparison, broadly investigated since several years [12], using several methods, including probabilistic [13] and possiblistic [14] ones, among others.

Interval comparison aims to rank real intervals. The heuristic approach developed in [11] has the advantage of not relying on midpoints for interval comparison. This makes sense particularly in the case of fuzzy numbers or confidence intervals.

The degree of preference $\Pi(A > B)$ of $A = [a_1, a_2]$ over $B = [b_1, b_2]$ is defined in [11] as:

$$\Pi(A > B) = \frac{\max(0, a_2 - b_1) - \max(0, a_1 - b_2)}{(a_2 - a_1) + (b_2 - b_1)}. \tag{4}$$

We observe that $P(A > B) + P(B > A) = 1$. Moreover,

$$\begin{cases} \text{if} \quad A \equiv B \quad \text{then} \quad \Pi(A > B) = \Pi(B > A) = 0.5, \\ \text{if} \quad a_2 < b_1 \quad \text{then} \quad \Pi(B > A) = 1. \end{cases} \tag{5}$$

We employ this comparison to rank class memberships $M_{i,j}$, $j \in \{1, C\}$. The defuzzification for the final decision simply consists in choosing the “least preferred” (minimum-error) one:

$$\text{Event}(i) = \arg \min_{j=1,\ldots,N} \{\Pi(M_{i,j} > M_{i,k \neq j})\}, \tag{6}$$

4. Experiments and results

4.1. Audio database

Different audio traffic datasets are suggested in the literature, such as AXA database [15], WASN [16] and MIVIA dataset [6]. The latter has the advantage to be the only open-access database for audio traffic surveillance. It contains nearly one hour of traffic sounds that were recorded in a real road environment at 23 locations in the province of Salerno, Italy, either in city center, highways or country roads. The database is segmented in 57 clips, of nearly one minute each, that were annotated manually. The annotation file includes the event labels, e.g. accident, tire skidding, horn blowing, etc., and the onset and offset times. Some audio events are considered as *Anomalous*, i.e. car crash, tire skidding and harsh braking, whereas all other events are considered as *Normal*, such as the sound of cars and pedestrians, and the background noise.

4.2. Parameter setting

The main parameter adjustment concerns the setting of the thresholds τ_j. Different values were experimentally optimized. Thresholds were pondered using the complementary of the proportion of each class as a weighting coefficient. Thus, the threshold τ_j for each class $j = 1, \ldots, N$ of each VAE's error was set as the baseline VAE's threshold τ_0 pondered by the weight $w_j = 1 - p_j$, where p_j is the proportion of samples of Class j. Table 1 summarizes the values.

Table 1
Parameter setting for the VAE's error and the fuzzy membership function (p_j is the proportion of Class j samples in the training set)

Part	Parameter	Value
All	Event weight w_j	$1 - p_j$
Baseline VAE	Error threshold τ_0	$\tau_0 \in]0, 1[$
Event-based VAE's	Error threshold τ_j	$\tau_0 \times w_j$

Table 2
Results of anomalous SED using VAE's and fuzzy membership function for *Normal* vs. *Anomalous* event classification ($p_{norm} = 0.79$ and $p_{anom} = 0.21$ are the proportions of *Normal* and *Anomalous* samples in the training set); For OC-SVM, the parameters $\nu =$ and γ are set to 0.14 and 2.5e-5, respectively, for their high performance.

Method	w_{norm}	w_{anom}	$Accuracy$	P_1	P_2	R_1	R_2	$F1_1$	$F1_2$
One-Class SVM			**0.84**	**0.94**	**0.59**	**0.86**	**0.77**	**0.90**	**0.67**
VAE with	0.5	0.5	0.83	0.95	0.38	0.86	0.65	0.90	0.48
fuzzy membership	**0.6**	**0.4**	**0.95**	**0.94**	**1.00**	**1.00**	**0.57**	**0.97**	**0.72**
	0.7	0.3	0.93	0.92	1.00	1.00	0.50	0.96	0.67
	0.8	0.2	0.93	0.92	1.00	1.00	0.40	0.96	0.57
	0.9	0.1	0.93	0.93	1.00	1.00	0.40	0.96	0.57

4.3. Experimental protocol

The experimental work aims to detect audio events on roads. To do so, features were extracted from the selected audio database, MIVIA DB [6], then experiments were realized following the steps described in Section 3.

Regarding the first step, i.e. feature extraction, data augmentation was realized to cope with the issue of rareness of *Anomalous* samples, so that more data is obtained through the segmentation of the audio signals into short frames, with a duration of 250 ms, with a high overlap rate, i.e. 75%.Nevertheless, it is worth noting that all training segments, whether belonging to *Normal* or *Anomalous*, contain background street noise.

Regarding neural networks training, the VAE network was constructed using convolutional layers, using an input feature vector made of log-energy and MFCC features, along with their first and second derivatives (Δ and Δ-Δ). 80% of the extracted data were utilized for training and validation, whereas test was realized on the remaining 20%.

4.4. Analysis of results

The evaluation results are listed in Table 2. These results correspond correspond to a state-of-the-art method, i.e. OC-SVM (used for benchmarking), and to the proposed method (event-based

VAE with fuzzy membership). For the latter, the values of the event weights were $\{w_j\}_{j=1,\ldots,N}$ were varied to find the tradeoff between data distribution and the global performance. For evaluation purposes, standard metrics were calculated, i.e. overall accuracy (Acc), precision (P), recall (R) and $F1$ scores, defined as in (7):

$$P_j = \frac{c_j}{e_j}, R_j = \frac{c_j}{r_j}, F1_j = \frac{2P_jR_j}{P_j + R_j}, \tag{7}$$

where r_j, e_j and c_j ($j \in \{1, 2\}$) are the number of ground-truth, estimated and correctly detected events for *Normal* and *Anomalous* class, respectively.

The results mentioned in Table 2 show the efficiency of using an interval-valued fuzzy membership function to improve anomaly detection. The main advantages of using such a method can be summarized as follows:

- The proposed methods outperforms the state-of-the-art OC-SVM, in terms of overall accuracy and balance between class-based metrics.
- Overall accuracy rates are enhanced, reaching 95% for the proposed method, vs. 84% for OC-SVM. Also, the precision, recall and F1 score obtained are more balanced between *Normal* and *Anomalous* classes, notwithstanding their disproportional distribution.
- The effect of using unbalanced weights is more evidenced, with higher accuracy when w_j is higher for the *Anomalous* class.

5. Discussion and conclusion

This paper presented a novel method of anomaly detection, based on interval-valued fuzzy sets. A direct application in road traffic surveillance allows detecting hazardous events such as car accidents using audio signals. The proposed method is based on combining two anomaly detection tools, i.e. auto-regressive VAE's and interval-valued fuzzy sets. Finally, a probabilistic interval comparison method, denoted as degree of preference, is utilized for defuzzification, i.e. detecting the corresponding class.

The main results can be summarized as follows: a) Spectrogram-extracted features are the most suitable to approach such a problem; b) unbalanced weights, where the least abundant class receives the highest weight, contribute to enhance the results; and c) interval-valued fuzzy sets seem more efficient than crisp one-class SVM to detect anomaly. As an outlook, the proposed method could be further improved in two directions: either by making it semi-supervised, as only normal data can be collected and trained, or fully unsupervised, by not using labels any more.

Acknowledgments

This work was carried out in the framework of the project *Xpert* funded by the University of Genova.

References

[1] A. A. Sodemann, M. P. Ross, B. J. Borghetti, A review of anomaly detection in automated surveillance, IEEE Transactions on Systems, Man, and Cybernetics, Part C 42 (2012) 1257–1272.

[2] J. M. Mendel, R. I. B. John, Type-2 fuzzy sets made simple, IEEE Transactions on Fuzzy Systems 10 (2002) 117–127. doi:10.1109/91.995115.

[3] P. Sevastianov, Numerical methods for interval and fuzzy number comparison based on the probabilistic approach and dempster–shafer theory, Information Sciences 177 (2007) 4645–4661.

[4] S. Ntalampiras, I. Potamitis, N. Fakotakis, Probabilistic novelty detection for acoustic surveillance under real-world conditions, IEEE Transactions on Multimedia 13 (2011) 713–719.

[5] T. Heittola, A. Mesaros, A. Eronen, T. Virtanen, Context-dependent sound event detection, EURASIP Journal on Audio, Speech, and Music Processing 2013 (2013) 1–13.

[6] P. Foggia, N. Petkov, A. Saggese, N. Strisciuglio, M. Vento, Audio surveillance of roads: A system for detecting anomalous sounds, IEEE transactions on intelligent transportation systems 17 (2015) 279–288.

[7] S. Rovetta, Z. Mnasri, F. Masulli, Detection of hazardous road events from audio streams: An ensemble outlier detection approach, in: 2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS), IEEE, 2020, pp. 1–6.

[8] Q. WEI, Y. LIU, Auto-encoder and metric-learning for anomalous sound detection task (2020). URL: http://dcase.community/challenge2020/index, preprint: http://dcase.community/documents/challenge2020/technical_reports/DCASE2020_Wei_49_t2.pdf.

[9] H. Purohit, R. Tanabe, T. Endo, K. Suefusa, Y. Nikaido, Y. Kawaguchi, Deep autoencoding gmm-based unsupervised anomaly detection in acoustic signals and its hyper-parameter optimization, in: Proceedings of the Detection and Classification of Acoustic Scenes and Events 2020 Workshop (DCASE2020), Tokyo, Japan, 2020.

[10] D. P. Kingma, M. Welling, Auto-encoding variational bayes, arXiv preprint 1312.6114 (2013).

[11] Y.-M. Wang, J.-B. Yang, D.-L. Xu, A preference aggregation method through the estimation of utility intervals, Computers & Operations Research 32 (2005) 2027–2049.

[12] E. Lee, R.-J. Li, Comparison of fuzzy numbers based on the probability measure of fuzzy events, Computers & Mathematics with Applications 15 (1988) 887–896.

[13] V.-N. Huynh, Y. Nakamori, J. Lawry, A probability-based approach to comparison of fuzzy numbers and applications to target-oriented decision making, IEEE Transactions on Fuzzy Systems 16 (2008) 371–387.

[14] A. Kasperski, A possibilistic approach to sequencing problems with fuzzy parameters, Fuzzy Sets and Systems 150 (2005) 77–86.

[15] M. Sammarco, M. Detyniecki, Crashzam: Sound-based car crash detection., in: Proceedings of Vehicle Technology and Intelligent Transport Systems (VEHITS), 2018, pp. 27–35.

[16] R. M. Alsina-Pagès, F. Orga, F. Alías, J. C. Socoró, A wasn-based suburban dataset for anomalous noise event detection on dynamic road-traffic noise mapping, Sensors 19 (2019) 2480.