

CADA-FVAE-GAN: ADVERSARIAL TRAINING FOR FEW-SHOT EVENT DETECTION

Xiaoxiang Zhu, Mengshu Hou, Xiaoyang Zeng and Hao Zhu

School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, China

ABSTRACT

Most supervised systems of event detection (ED) task rely heavily on manual annotations and suffer from high-cost human effort when applied to new event types. To tackle this general problem, we turn our attention to few-shot learning (FSL). As a typical solution to FSL, cross-modal feature generation based frameworks achieve promising performance on images classification, which inspires us to advance this approach to ED task. In this work, we propose a model which extracts latent semantic features from event mentions, type structures and type names, then these three modalities are mapped into a shared low-dimension latent space by modality-specific aligned variational autoencoder enhanced by adversarial training. We evaluate the quality of our latent representations by training a CNN classifier to perform ED task. Experiments conducted on ACE2005 dataset show an improvement with 12.67% on F1-score when introducing adversarial training to VAE model, and our method is comparable with existing transfer learning framework for ED.

KEYWORDS

Event Detection, Few-Shot Learning, Cross-modal generation, Variational autoencoder, GAN

1. INTRODUCTION

As an essential subtask for IR, ED aims at identifying the event triggers in the text and assigning the pre-defined event types to each of the triggers. There are 33 types of events according to the ACE2005 corpus, such as “Attack”, “Transport”, “Die” etc. For instance, in the sentence “Tuesday’s southern Philippines airport blast”, “blast” is the trigger of event “Attack”, ED system should identify the word “blast” and categorize it to the corresponding event type.

ED task is usually modeled as the multi-classification problem in the traditional supervised methods. These methods suffer from the heavy reliance on manual annotations and features specific to the particular event types, which makes it difficult to handle new or unseen types without additional human annotation efforts. In order to overcome this challenge, we model ED task with transfer learning approaches.

Few-shot learning framework, as a typical solution for transfer learning, which enables models to handle classification task for new classes of examples, gives us a valuable inspiration. The goal of FSL is to learn transferable knowledge from training examples (*seen* classes) to test examples (*unseen* classes), with only a few examples moved from test into training examples. Zero-shot learning (ZSL) is another framework similar to FSL, where the classes of training and test

examples are absolutely disjoint. Both zero- and few-shot learning approaches typically exploit semantic knowledge to achieve transferability. [1-4] improves zero-shot predictions of images with semantic knowledge learned from unconstructed text description. Neural Snowball [5] is a few-shot relation extraction (RE) framework transferring semantic knowledge from existing relations to new ones. [6] designs a hybrid attention-based neural model to improve noisy few-shot relation classification (RC) by grasping external knowledge. [7] applies ZSL to event extraction problem by learning a generic mapping function of event types and mapping both event mentions (trigger and context) and types into a shared semantic space.

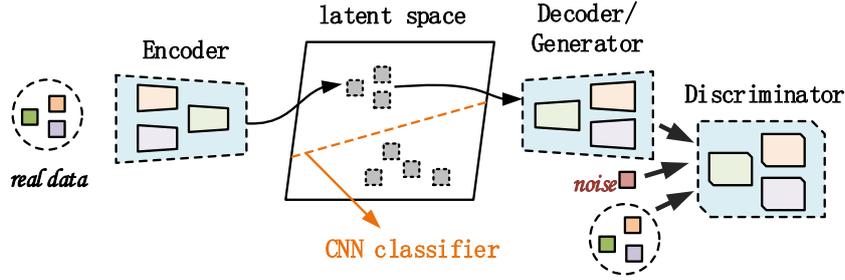


Figure 1. A review of our CADA-fVAE-GAN model

Our CADA-fVAE-GAN is a cross-modal framework utilizing VAE as its key module for feature generation. Recently, cross-modal deep learning model has received much attention. [8] proposes a ZSL classification framework, on which image features and its descriptive text of categories are mapped into a shared semantic word vector space. [9] and [10] jointly learn multi-modal representations with distribution alignment in their latent space. [11] uses VAE-based generative model to perform generalized ZSL/FSL via images classification problem, by mapping multi-modal samples into a shared latent low-dimension feature space, which achieves encouraging results. Significantly, [11] indicates that latent features constructed by VAE are semantically interpretable for classification. Therefore, we decide to advance the model to few-shot learning for event detection framework, and study the generality of latent features inferred by VAE encoder. The superiority of VAE lies in variational inference, while its generation performance is not comparable to some powerful generative models such as GAN. In some degree, this is partly due to the fact that VAE is not capable to encode high-quality latent representations. To alleviate this problem, [12] develops a conditional generative model with the combination of VAE and GAN, which learns highly discriminative features for downstream task. [13] introduces adversarial training to VAE for better variational inference. [14] combines VAE with GAN and utilizes learned features in data space for better measurement of similarities.

Inspired by adversarial training of GAN and its attractive usage in NLP [15-18], we fold the generator of GAN and the decoder of VAE into one, realizing the sharing of neural parameters and training process. We notice that the input for our model consists of three modalities, including event mention structures parsed by AMR, event type structures and type name embeddings, which are not considered as high-level representations. And experimental results indicate that using abstract features extracted from these modalities by CNN will degrade the interpretability of latent representations constructed by VAE encoder. However, a CNN classifier is able to extract valuable features from latent representations with original modalities as input, and produce acceptable classification results for our ED task. Figure 1 is a brief review of our CADA-fVAE-GAN model. In summary, our main contributions is three-fold:

1. We apply VAE-based generative model to few-shot learning of event detection for the first time, and demonstrate the transferability of latent representations constructed by VAE.
2. We combine VAE with GAN to improve the quality of latent representations and the transferability of the model via adversarial training.
3. Experiments conducted on ACE2005 dataset achieve ideal results, which demonstrates the effectiveness of our proposed model, and indicates some promising direction for further research on ED problem.

Next, in Section 2, we discuss several representative works on event detection, including traditional feature based methods and recent neural network based methods. In Section 3, we explore the architecture of our CADA-fVAE-GAN model, and then we perform ablation study to evaluate each module in Section 4. Finally, Section 5 concludes our work and discusses possible future scope for further research.

2. LITERATURE REVIEW

Recently, much more attention has been attracted on event extraction. Traditional methods are mainly based on feature-learning. [22-25] rely on fine-annotated textual features to identify the types of event triggers. With the emergence of deep neural network, [26-28] exploit convolution neural network (CNN) to construct higher-granularity informative representations through stacked convolution layers, which prove the feasibility of CNN on event detection. Moreover, as a typical sequential model, recurrent neural network (RNN) is equipped with the qualities to perform sentence classification task. [29] extracts syntactic relations by constructing dependency bridges over Bi-LSTM. [30] introduces document-level information to bidirectional RNN and alleviating the complexity for inference rules. [31] builds document embeddings and supervised attention to enhance event trigger identification and classification. [32] combines CNN and Bi-LSTM to extract informative representations for event detection. Recent large pre-trained language model such as BERT [33] and ELMo [34] also attract some researchers: [35] using a transition-based framework and BERT embeddings, [17] using BERT based encoders and adversarial training mechanism, [36] using a Bi-LSTM with BERT token representations, [16] introducing an incremental learning framework with ELMo word representations.

3. METHODOLOGY

We model event detection as a multi-classification task. The inputs for our model include three modalities: event mention structures constructed by AMR and their corresponding type structures and type name embeddings pre-defined in ACE2005 corpus. So we give three VAEs, one for each modality. To improve reconstruction quality of VAE, we collapse decoder in VAE and generator in GAN into one, by sharing neural network parameters and training process. In our model, decoders not only need to reconstruct low-dimension latent representations, which is the basic function of VAE, but also play a role of generator in GAN. Therefore, a randomized noise is constructed as additional input for each decoder/generator. Eventually, $M+1$ (M is the number of modalities, namely 3 in this work) outputs are produced from each decoder/generator, which will be treated as *fake* data for the input of discriminator. And the original input for VAE encoder will be fed into discriminator as *real* data. Discriminator takes its responsibility of differentiating *fake* data and *real* data. The detailed architecture of the model is shown as Figure 2, and a brief training procedure is listed in Algorithm 1. Intuitively, due to the introduction of GAN, VAE decoder will be improved to produce higher-quality reconstructions after adversarial training process. In this way, VAE encoder is expected to construct latent features with richer semantic meanings, which is beneficial to downstream task.

3.1. AMR Semantic Graph Encoder

FSL usually needs high-quality class representations to learn a robust transfer learning model. Following [7], we take advantage of AMR to build semantic graph by identifying event triggers and arguments (such as *Time*, *Location*, *Person*, etc). For instance, the AMR-parsed event mention structure of the sentence “1994 civil war in Rwanda, where government-led militia slaughtered an estimated 800,000 opposition,...” and “Toefing transferred to Bolton in February

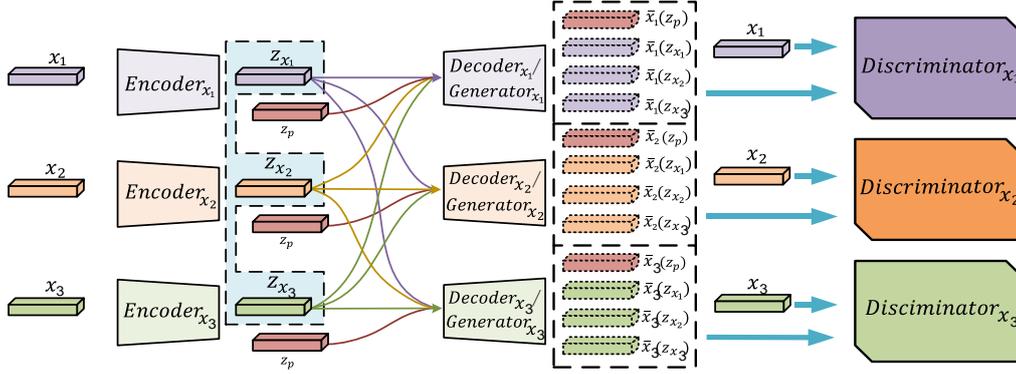


Figure 2. Architecture CADA-fVAE-GAN model

Algorithm 1 Training Procedure

Input:

- $M = 3$ is the number of modalities
 - $N_{epoch} = 80$ is the number of training epoch
 - N_d is the iterations to train discriminator in each epoch
 - $\omega = 1000$ is the weighting factor for \mathcal{L}_G
 - 1: **for** epoch in $1, 2, \dots, N_{epoch}$ **do**
 - 2: Sample x_1, x_2, \dots, x_M from M modalities
 - 3: **# Train discriminator**
 - 4: **for** n in $1, 2, \dots, N_d$ **do**
 - 5: Use VAE encoder to map x_1, x_2, \dots, x_M into latent space as $z_{x_1}, z_{x_2}, \dots, z_{x_M}$
 - 6: Sample a random noise z_p from $\mathcal{N}(0, I)$
 - 7: Use VAE decoder to produce reconstructions $\bar{x}_1(z_{x_1}), \bar{x}_2(z_{x_2}), \dots, \bar{x}_M(z_{x_i}), \bar{x}_i(z_p)$ for all i in $1, 2, \dots, M$
 - 8: Compute \mathcal{L}_D with $\bar{x}_1(z_{x_1}), \bar{x}_2(z_{x_2}), \dots, \bar{x}_M(z_{x_i}), \bar{x}_i(z_p)$ and x_i
 - 9: $w \leftarrow Adam(\nabla \mathcal{L}_D, w)$
 - 10: **end for**
 - 11: **# Train VAE like a generator**
 - 12: Use VAE encoder to map x_1, x_2, \dots, x_M into latent space as $z_{x_1}, z_{x_2}, \dots, z_{x_M}$
 - 13: Use VAE decoder to produce reconstructions $\bar{x}_1(z_{x_1}), \bar{x}_2(z_{x_2}), \dots, \bar{x}_M(z_{x_i}), \bar{x}_i(z_p)$ for all i in $1, 2, \dots, M$
 - 14: Compute $\mathcal{L}_{fVAE}, \mathcal{L}_{CA}, \mathcal{L}_{DA}$
 - 15: Use decoder as generator to compute \mathcal{L}_G
 - 16: $w \leftarrow Adam(\nabla \mathcal{L}_{CADA-fVAE} + \omega \mathcal{L}_G, w)$
 - 17: **end for**
-

2002 from German club Hamburg.” are shown in the Figure 3 (top). Figure 3 (bottom) also shows their pre-defined event type structures in ACE2005 dataset. Considering the shared semantic meaning between an event trigger and its type name, and the similarity between mention structure and its type structure, we exploit multi-modal VAEs to map three modalities into a shared latent space, then extract their semantic representations.

3.2. Preprocessing for Multi-Modal Structural Features

According to the learned event mention structures, we represent each edge in the directed graph as a tuple $u = \langle w_1, \epsilon, w_2 \rangle$, where w_1, w_2 denote word entities at endpoints, ϵ denote the AMR relation between w_1 and w_2 , such as $\langle \text{war}, \text{:mod}, \text{civil} \rangle$. For each event mention structure, we fix the number of binary relations to r , then map w_1 and w_2 to their word embeddings V_{w_1} and $V_{w_2} \in R^d$, where d is dimension of word embeddings. Then V_{w_1} and V_{w_2} are concatenated and we can get a matrix $M_u \in R^{2d \times r}$ representing all the relations in the event mention. Assume that $M_\epsilon \in R^{2d \times 2d}$ is the matrix representation of AMR relation ϵ , then $M_\epsilon \times M_u$ is the composition representation for the event mention structure.

For the type structure, each edge in the graph is represented as $v = \langle a, b \rangle$. The number of such tuple in each type structure is also fixed to r . Concatenate embeddings of word entities a and b , namely V_a and $V_b \in R^d$, we get the matrix representation of type structure $M_v \in R^{2d \times r}$.

As for the event type name denoted by t , we simply use its word embedding $V_t \in R^d$.

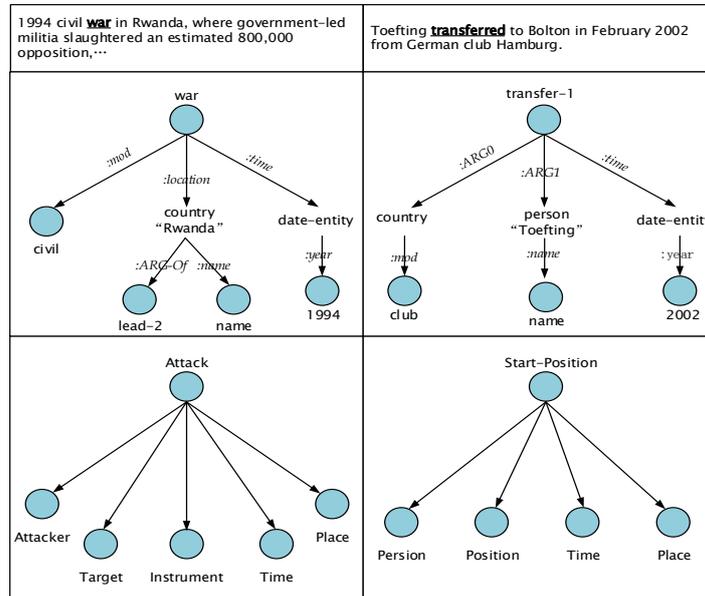


Figure 3. Examples of Event Mention Structure and Type Structures

In AMR graph, each edge with the keyword “:<arg-name>” represents the semantic relation between two word entities or concepts, including affiliation, coreference, category definition and target orientation. The root node is usually the central word in a sentence, such as event name or an action name

3.3. Adversarial Latent Features Generating Network

In this section, we propose a VAE-based feature generation model with cross-distribution-aligned and adversarial training for event detection, which is called CADA-fVAE-GAN. The model exploits adversarial training to strengthen VAE and few-shot learning process.

3.3.1. Basic VAE

Variational autoencoder (VAE) is a typical generative neural network consisting of an encoder and a decoder. VAE encoder maps the given data into a latent feature space, and the decoder reconstruct latent features and maps them back to original data space. Different from trivial autoencoder, VAE is skilled at inferring the true conditional probability distribution of latent variables $z \sim p_\theta(z|x)$. VAE performs this by approximating a closest posterior distribution $q_\phi(z|x) \sim \mathcal{N}(\mu, \Sigma)$, and minimize their variational distance. The objective function of trivial VAE is written as:

$$\mathcal{L}(\theta, \phi; x) = \mathbb{E}_{z \sim q_\phi(z|x)} [\log p_\theta(x|z)] - D_{KL}[q_\phi(z|x) \parallel p_\theta(z)]$$

where the first RHS term is reconstruction loss, the second term is Kullback-Leibler divergence (KLD) between $q_\phi(z|x)$ and $p_\theta(z)$, which can be written as followed in Gaussian case:

$$D_{KL}[q_\phi(z|x) \parallel p_\theta(z)] = -\frac{1}{2} \sum_{j=1}^J (1 + \log \sigma_j^2 - \mu_j^2 - \sigma_j^2)$$

where J is dimensionality of z , μ_j and σ_j denote each element of mean and s.d. evaluated at datapoint j .

In addition, when reconstructing original samples from latent variables z , we can adopt a reparameterization trick as followed:

$$\tilde{z} = \mu_z + \sigma_z \odot \varepsilon, \quad \text{where } \varepsilon \sim \mathcal{N}(0, I)$$

3.3.2. Cross- and Distribution-Aligned VAE

In our few-shot event detection framework, for unseen classes, only category descriptions and a few mention samples are provided to training set. Therefore, it is necessary for the model to have the capability of cross-modal generalization. Namely, one modality-specific encoder/decoder is expected to encode/decode another modalities with high-quality. For the better performance, we exploit β -VAE [19]. Since each modality have its specific VAE, such that x_1 for event mention structure, x_2 for type structure and x_3 for type name embedding, so the final loss for our basic fVAE is:

$$\mathcal{L}_{fVAE} = \sum_i^M \mathbb{E}_{z \sim q_\phi(z|x)} [\log p_\theta(x^{(i)}|z^{(i)})] - \beta D_{KL}[q_\phi(z^{(i)}|x^{(i)}) \parallel p_\theta(z^{(i)})]$$

By weighting KLD with β , we can produce better reconstructions than trivial VAE. Now we introduce constraint to cross-modal reconstruction for every modality-specific VAE. As is depicted in Figure 2, each decoder should learn to reconstruct latent representations from other $M - 1$ modalities, which leads to our cross-aligned loss:

$$\mathcal{L}_{CA} = \sum_i^M \sum_{j \neq i}^M \|x^{(i)} - DEC_j(E_i(x^{(i)}))\|^2$$

where E_i and DEC_j denote objective functions of encoder for modality i , and decoder for modality j , respectively.

Furthermore, distributions of different latent variables are aligned by minimizing their Wasserstein distance [20]. The Wasserstein distance between two Gaussian distributions is:

$$W_{i,j} = [\|\mu_i - \mu_j\|^2 + tr(\Sigma_i) + tr(\Sigma_j) - 2(\Sigma_i^{\frac{1}{2}}\Sigma_i\Sigma_j^{\frac{1}{2}})^{\frac{1}{2}}]^{\frac{1}{2}}$$

Since covariance matrices constructed by encoder are diagonal and commutative, we can simplify this equation to:

$$W_{i,j} = (\|\mu_i - \mu_j\|^2 + \|\Sigma_i - \Sigma_j\|^2)^{\frac{1}{2}}$$

So the total loss of distribution-aligned for M modalities is:

$$\mathcal{L}_{DA} = \sum_i^M \sum_{j \neq i}^M W_{i,j}$$

We combine the basic VAE loss \mathcal{L}_{fVAE} with cross- and distribution-aligned:

$$\mathcal{L}_{CADA-fVAE} = \mathcal{L}_{fVAE} + \zeta \mathcal{L}_{CA} + \vartheta \mathcal{L}_{DA}$$

where ζ and ϑ respectively weight cross- and distribution-aligned loss.

3.3.3. Adversarial Training

Our model aims at providing an enlightening perspective to the semantic representation of latent features via a classification task in the NLP field. Intuitively, the higher quality of reconstructions indicates the more interpretable latent representations. Under the constraint of VAE objective function, improving decoder will accordingly improve encoder. Moreover, it has been shown that combining VAE and GAN leads to better generation results [12-14]. Inspired by the superiority of adversarial training strategy, we decide to link a discriminator following VAE decoder, and decoder plays a role of generator in GAN.

WGAN [21] has been proved to have better theoretical properties than the vanilla GAN, for which we choose WGAN in our model. According to the architecture shown as Figure 2, the losses of generators and discriminators are:

$$\begin{aligned} \mathcal{L}_G &= \sum_i^M \mathbb{E}[D_i(G_i(z_p^{(i)}))] + \sum_i^M \sum_j^M \mathbb{E}[D_i(G_i(z_{x(j)}))] \\ \mathcal{L}_D &= \sum_i^M \mathbb{E}[D_i(x^{(i)})] - \sum_i^M \mathbb{E}[D_i(G_i(z_p^{(i)}))] - \sum_i^M \sum_j^M \mathbb{E}[D_i(G_i(z_{x(j)}))] - \sum_i^M \lambda_{G_i} \end{aligned}$$

where G_i and D_i are generator and discriminator for modality i , $z_p^{(i)}$ is random noise sampled from $\mathcal{N}(0, I)$ for modality i , and $z_{x^{(i)}}$ is latent representation for modality i . \mathbb{G}_i is gradient penalty for modality i , with a penalty coefficient λ :

$$\mathbb{G}_i = \mathbb{E} \left[\left(\left\| \nabla_{\hat{x}_{z_p^{(i)}}} D_i(\hat{x}_{z_p^{(i)}}) \right\|_2 \right)^2 \right] + \sum_j^M \mathbb{E} [(\left\| \nabla_{\hat{x}^{(j)}} D_i(\hat{x}^{(j)}) \right\|_2)^2]$$

where $\hat{x}_{z_p^{(i)}} = x^{(i)} + \alpha(\tilde{x}_{z_p^{(i)}} - x^{(i)})$, $\hat{x}^{(j)} = x^{(j)} + \alpha(\tilde{x}^{(j)} - x^{(j)})$ with $\alpha \sim U(0,1)$, $x^{(i)}$ and $x^{(j)}$ are the real sample for modality i and j respectively, $\tilde{x}_{z_p^{(i)}}$ is reconstructed from random noise $z_p^{(i)}$, $\tilde{x}^{(j)}$ is reconstruction for modality j .

Final objective function is:

$$\min_{CADA-fVAE,G} \max_D \mathcal{L}_{CADA-fVAE} + \omega \mathcal{L}_G + \mathcal{L}_D$$

where ω is the weighting factor.

3.3.4. Implementation details

All encoders and decoders are implemented as MLPs with one hidden layer, which will not degrade performance. On the one hand, AMR graph abstractly represents event mention. On the other hand, a CNN classifier is used to predict event types, and higher-level semantic representations will be obtained further. More hidden layers lose key information. We find that 1560 hidden units for event mention structure encoders and 1660 for decoders produce better results in our work. The encoder of type name embeddings and type structures have 1450 hidden units and 665 for decoders.

The dimension of VAE latent space is 120. Each discriminator is implemented as MLP with one hidden layer and 1450 units, whose output is activated by a Sigmoid. Following [11], gradient penalty coefficient λ is set to 10. We find that $\omega = 1000$ works well on ACE2005 dataset. During the training of 80 epochs, ζ is increased from epoch 6 to epoch 22 by a rate of 0.54 per epoch, while ϑ is increased from epoch 21 to 75 by 0.044 per epoch. As is suggested by [21], we update decoder/generator every 5 discriminator iterations. All modules including classifier are trained using Adam optimizers, with learning rate = 1.5e-4 for VAE and 5e-5 for discriminators. CNN classifier is trained for one epoch with learning rate=1e-3 and CrossEntropyLoss as its criterion. CNN classifier is implemented with two one-dimension convolution layers, each of which contains a ReLU and a MaxPool1d. Final predictions are produced by a fully connected layer.

4. EXPERIMENTS

4.1. Settings

ACE2005 dataset defines 33 event types, on which experiments are conducted to evaluate the performance of the model. Training set contains the top-10 most popular event types (*Attack, Transport, Die, Meet, Sentence, Arrest-Jail, Transfer-Money, Elect, Transfer-Ownership, End-Position*) as seen types, and the remaining 23 types are selected as unseen types, which are included in the test set. In the few-shot learning, n examples of event mention features per class

are moved from the test to the training set, where n is set to 2. We use P (Precision), R (Recall), F1-score and H (Harmonic mean) as performance metrics. Note that:

$$P = \frac{TP}{TP + FP}, \quad R = \frac{TP}{TP + FN}, \quad F1 = \frac{2PR}{P + R}$$

When evaluating the quality of a multi-classification task, TP is # of true positives, FP is # of false positives and FN is # of false negatives.

$$H = \mathbb{E} \left(\sum_i acc_i \right),$$

where acc_i is the accuracy that samples of i^{th} event type are predicated correctly in the given sequence.

4.2. Ablation Study

In this section, we analyze crucial building modules in the proposed model by disabling each of them, respectively.

fVAE is the baseline only using β -VAE without cross-aligned, distribution-aligned and adversarial training.

CA-fVAE, DA-fVAE, CADA-fVAE are the baseline models using β -VAE with cross-aligned, distribution-aligned and both of them.

CADA-fVAE-GAN combines WGAN with CADA-fVAE to improve the quality of latent representations constructed by VAE framework.

We can draw conclusions from Table 1 that both cross-aligned and distribution-aligned improve the performance. The cross-alignment works better than distribution-alignment (36.21% vs. 31.59% on F1-score, 49.52% vs. 46.39% on H), and more outstanding results are produced by compositing two tricks. Moreover, the introduction of GAN further improves the performance. Compared with CADA-fVAE, our CADA-fVAE-GAN increases the test results by 0.13% on P, 6.78% on R, 12.67% on F1 and 2.38% on H. Ablation study shows the adversarial training leads the encoder of VAE to producing higher-quality latent representations by improving the VAE decoder directly, under the restraint of VAE objective function.

Table 1. Results of ablation study.

	P (%)	R (%)	F1 (%)	H (%)
fVAE	31.06	40.08	24.96	40.10
CA-fVAE	39.92	48.26	36.21	49.52
DA-fVAE	39.58	45.72	31.59	46.39
CADA-fVAE	42.90	45.54	37.84	52.04
CADA-fVAE-GAN	43.21	52.36	50.51	54.42

4.3. Model Comparison

In this section, we show that the proposed few-shot learning model achieves comparable performance with existing transfer learning framework for ED. We compare our method with the following baseline:

Transfer: [7] design a transferable architecture for event extraction, using CNN to generate vector representations for the event mention and event type structure. The top-10 most popular event types in ACE2005 chosen by us as *seen* types are the same as [7]. Table 2 shows the performance.

Table 2. Event trigger classification performance on unseen ACE2005 event types.

	P (%)	R (%)	F1 (%)	H (%)
Transfer	75.50	36.30	49.10	-
CADA-fVAE-GAN	43.21	52.36	50.51	54.42

CADA-fVAE-GAN exploits latent representations encoded by VAE, which is proved to be comparable with CNN representations generated by **Transfer**. Conclusions can be drawn from the above results that VAE+GAN could be used to generate features for ED task in transfer learning situations.

5. CONCLUSIONS

In this paper we propose a few-shot event detection model named CADA-fVAE-GAN, which introduces adversarial training to variational autoencoders (VAE). To improve the performance of VAE, cross- and distribution alignment are exploited. With cross-aligned latent distributions and reconstructions, latent representations are enriched by more interpretable semantic meaning. Moreover, adversarial training provided by WGAN strengthens VAE encoder indirectly. Experiments conducted on ACE2005 dataset demonstrate the transferability of low-dimension latent semantic knowledge constructed by VAE and the effectiveness of adversarial training.

Future scope of the research is suggested to be focused on generalization improvements. Specifically, few-shot learning of event detection can be advanced to zero-shot, generalized few-shot and generalized zero-shot, which are of more practical value

6. ACKNOWLEDGEMENT

This research is supported by the grants from the National Key Research and Development Program of China (No. 2019YFB1705601)

REFERENCES

- [1] Frome, Andrea, et al. "DeViSE: A Deep Visual-Semantic Embedding Model." neural information processing systems (2013): 2121-2129.
- [2] Lampert, Christoph H., Hannes Nickisch, and Stefan Harmeling. "Learning to detect unseen object classes by between-class attribute transfer." computer vision and pattern recognition (2009): 951-958.
- [3] Reed, Scott, et al. "Learning Deep Representations of Fine-Grained Visual Descriptions." computer vision and pattern recognition (2016): 49-58.
- [4] Zhang, Li, Tao Xiang, and Shaogang Gong. "Learning a Deep Embedding Model for Zero-Shot Learning." computer vision and pattern recognition (2017): 3010-3019.
- [5] Gao, T., Han, X., Xie, R., Liu, Z., Lin, F., Lin, L., & Sun, M. (2020). Neural Snowball for Few-Shot Relation Learning. In AAI (pp. 7772-7779).
- [6] Gao, Tianyu, et al. "Hybrid Attention-based Prototypical Networks for Noisy Few-Shot Relation Classification." national conference on artificial intelligence (2019): 6407-6414.

- [7] Huang, Lifu, et al. "Zero-shot transfer learning for event extraction." arXiv preprint arXiv:1707.01066 (2017).
- [8] Socher, Richard, et al. "Zero-Shot Learning Through Cross-Modal Transfer." neural information processing systems (2013): 935-943.
- [9] Tsai, Yaohung Hubert, Liangkang Huang, and Ruslan Salakhutdinov. "Learning Robust Visual-Semantic Embeddings." international conference on computer vision (2017): 3591-3600.
- [10] Mukherjee, Tanmoy, Makoto Yamada, and Timothy M. Hospedales. "Deep matching autoencoders." arXiv preprint arXiv:1711.06047 (2017).
- [11] Schonfeld, Edgar, et al. "Generalized Zero- and Few-Shot Learning via Aligned Variational Autoencoders." computer vision and pattern recognition (2019): 8247-8255.
- [12] Xian, Yongqin, et al. "F-VAEGAN-D2: A Feature Generating Framework for Any-Shot Learning." computer vision and pattern recognition (2019): 10275-10284.
- [13] Yu, Xianwen, et al. "VAEGAN: A Collaborative Filtering Framework based on Adversarial Variational Autoencoders.." international joint conference on artificial intelligence (2019): 4206-4212.
- [14] Larsen, A. B. L., Sønderby, S. K., Larochelle, H., & Winther, O. (2016, June). Autoencoding beyond pixels using a learned similarity metric. In International conference on machine learning (pp. 1558-1566).
- [15] Zhu, Yizhe, et al. "A Generative Adversarial Approach for Zero-Shot Learning from Noisy Texts." computer vision and pattern recognition (2018): 1004-1013.
- [16] Lu, Yaojie, et al. "Distilling Discrimination and Generalization Knowledge for Event Detection via Delta-Representation Learning." meeting of the association for computational linguistics (2019): 4366-4376.
- [17] Wang, Xiaozhi, et al. "Adversarial Training for Weakly Supervised Event Detection." north american chapter of the association for computational linguistics (2019): 998-1008.
- [18] Hong, Yu, et al. "Self-regulation: Employing a Generative Adversarial Network to Improve Event Detection." meeting of the association for computational linguistics (2018): 515-526.
- [19] Higgins, I., Matthey, L., Pal, A., Burgess, C. P., Glorot, X., Botvinick, M., ... & Lerchner, A. (2017). beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework. international conference on learning representations.
- [20] Givens, Clark R., and Rae Michael Shortt. "A class of Wasserstein metrics for probability distributions." Michigan Mathematical Journal 31.2 (1984): 231-240
- [21] Adler, Jonas, and Sebastian Lunz. "Banach Wasserstein GAN." neural information processing systems (2018): 6755-6764.
- [22] Ahn, D. (2006, July). The stages of event extraction. In Proceedings of the Workshop on Annotating and Reasoning about Time and Events (pp. 1-8).
- [23] Ji, Heng, and Ralph Grishman. "Refining Event Extraction through Cross-Document Inference." meeting of the association for computational linguistics (2008): 254-262.
- [24] Gupta, Prashant, and Heng Ji. "Predicting Unknown Time Arguments based on Cross-Event Propagation." meeting of the association for computational linguistics (2009): 369-372.
- [25] Riedel, Sebastian, Limin Yao, and Andrew McCallum. "Modeling relations and their mentions without labeled text." european conference on machine learning (2010): 148-163.
- [26] Nguyen, Thien Huu, and Ralph Grishman. "Event Detection and Domain Adaptation with Convolutional Neural Networks." international joint conference on natural language processing (2015): 365-371.
- [27] Chen, Yubo, et al. "Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks." international joint conference on natural language processing (2015): 167-176.
- [28] Nguyen, Thien Huu, and Ralph Grishman. "Modeling Skip-Grams for Event Detection with Convolutional Neural Networks." empirical methods in natural language processing (2016): 886-891.
- [29] Sha, Lei, et al. "Jointly extracting event triggers and arguments by dependency-bridge rnn and tensor-based argument interaction." Thirty-Second AAAI Conference on Artificial Intelligence. 2018.
- [30] Duan, Shaoyang, Ruifang He, and Wenli Zhao. "Exploiting Document Level Information to Improve Event Detection via Recurrent Neural Networks." international joint conference on natural language processing (2017): 352-361.
- [31] Zhao, Yue, et al. "Document Embedding Enhanced Event Detection with Hierarchical and Supervised Attention." meeting of the association for computational linguistics (2018): 414-419.

- [32] Liu, Yingchi, et al. "Document Information Assisted Event Trigger Detection." international conference on big data (2018): 5383-5385.
- [33] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
- [34] Peters, Matthew E., et al. "DEEP CONTEXTUALIZED WORD REPRESENTATIONS." north american chapter of the association for computational linguistics (2018): 2227-2237.
- [35] Zhang, Junchi, et al. "Extracting Entities and Events as a Single Task Using a Transition-Based Neural Model." international joint conference on artificial intelligence (2019): 5422-5428.
- [36] Sims, Matthew, Jong Ho Park, and David Bamman. "Literary Event Detection." meeting of the association for computational linguistics (2019): 3623-3634.

AUTHORS

Xiaoxiang Zhu is Master's Degree in reading, at the Department of computer science and Engineering, University of Electronic Science and Technology (UESTC), Chengdu, China. He received the B.S. degrees in UESTC in 2018. His interests focus on distributed storage and computing.



Mengshu Hou is a professor in the school of computer science & engineering at the University of Electronic Science and Technology of China (UESTC). He received the M.S and Ph.D degrees in 2002 and 2005 respectively from the UESTC. His current research interests include distributed database management system, machine learning applied to natural language processing, and big data analytics.



Xiaoyang Zeng is currently a Ph.D. at the Department of computer science and Engineering, University of Electronic Science and Technology (UESTC), Chengdu, China. He received the B.S. degrees in Southwest Petroleum University in 2018, and passed the successive master-doctor program and is studying in UESTC. His research interests focus on natural language processing and text mining.



Hao Zhu is an engineer in the Information Center at the University of Electronic Science and Technology of China (UESTC). He received the B.S and M.S degrees in 2002 and 2006 respectively from the UESTC. His current research interests include management informatization, data visualization, and big data analysis

