ICPR 2020

Feature-aware unsupervised learning with joint variational attention and automatic clustering

Ru Wang, Lin Li, Peipei Wang, Xiaohui Tao, Peiyu Liu

Shandong Normal university

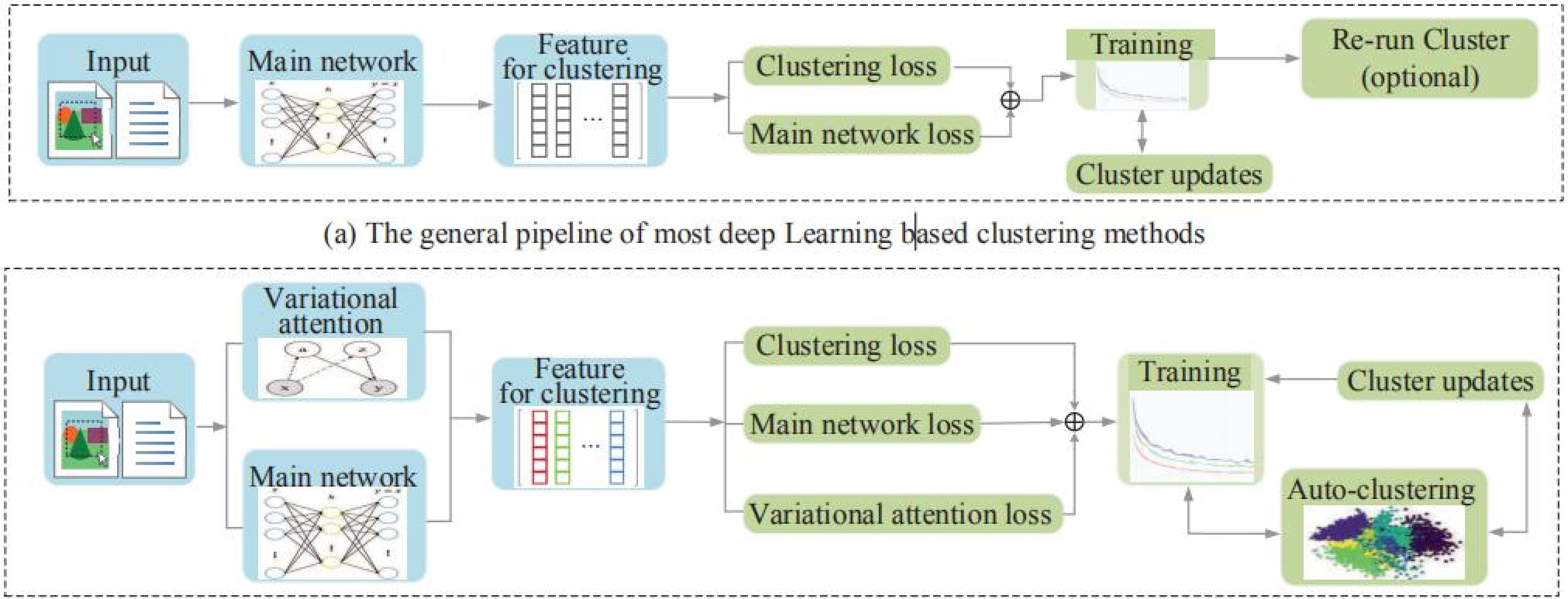
Wuhan University of Technology

University of Southern Queensland

Contents

Background

- Deep Clustering (DC) , a deep neural networks (DNNs)-based clustering method has become one of the effective ways to solve high-dimensional data categorization.
- However, the traditional attention mechanism is not suitable for deep clustering tasks, which brings many problems.



(b) Our unsupervised deep clustering pipeline based on variational attention and auto-clustering

Background

Main challenges

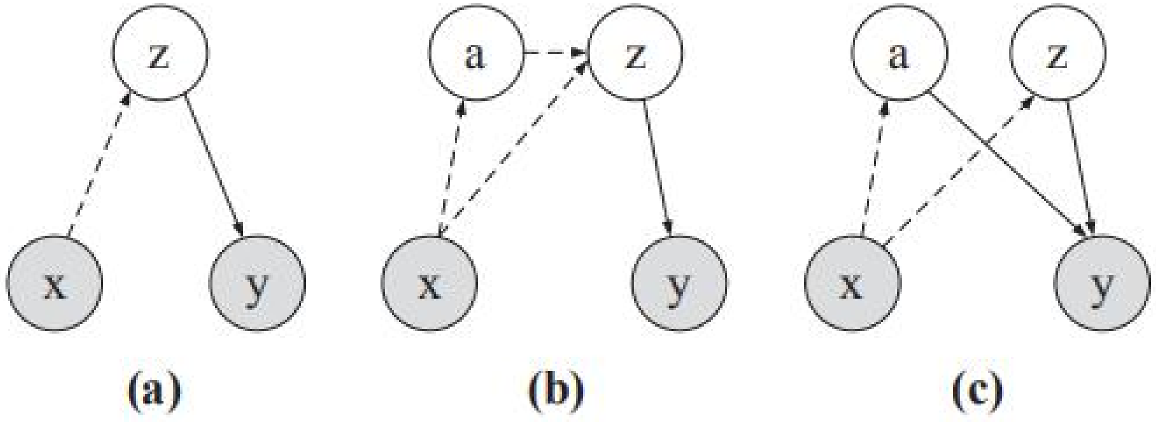
- The latent feature (z) maybe lose effectiveness when it utilizes traditional attention mechanism directly to process image and text data. This phenomenon can be understood that when hidden layer feature (z) are combined with attention vectors (a). This will cause the loss of effectiveness of the hidden layer feature (z).
- Feature adaptation of multi-stage model.

Graphical model representations.

(a) variational auto-encoder (VAE).

(b) variational encoder-decoder with traditional attention.

(c) Our variational attention encoder-decoder.



Background

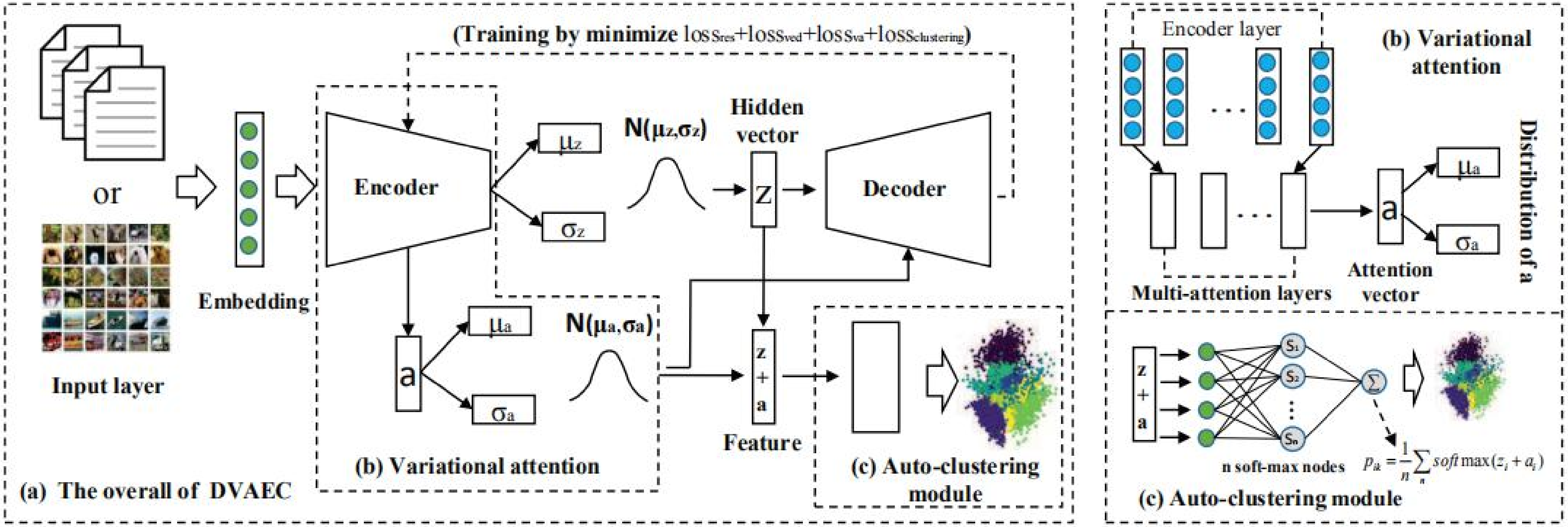
Our contributions:

- We propose a novel deep variational attention encoder-decoder for clustering that en hances implicit feature representation learning ability.
- We are the first to apply variational attention to deep clustering tasks.In order to avoi d the unreliability of sample similarity calculation in clustering task, we devise a feat ure-aware automatic clustering module to guide network learning and predict sample category.
- To improve the training process of the deep cluster ing task, we present an optimizati on objective approachwhich joints feature representation learning and clustering sim ultaneously.

Contents

Method

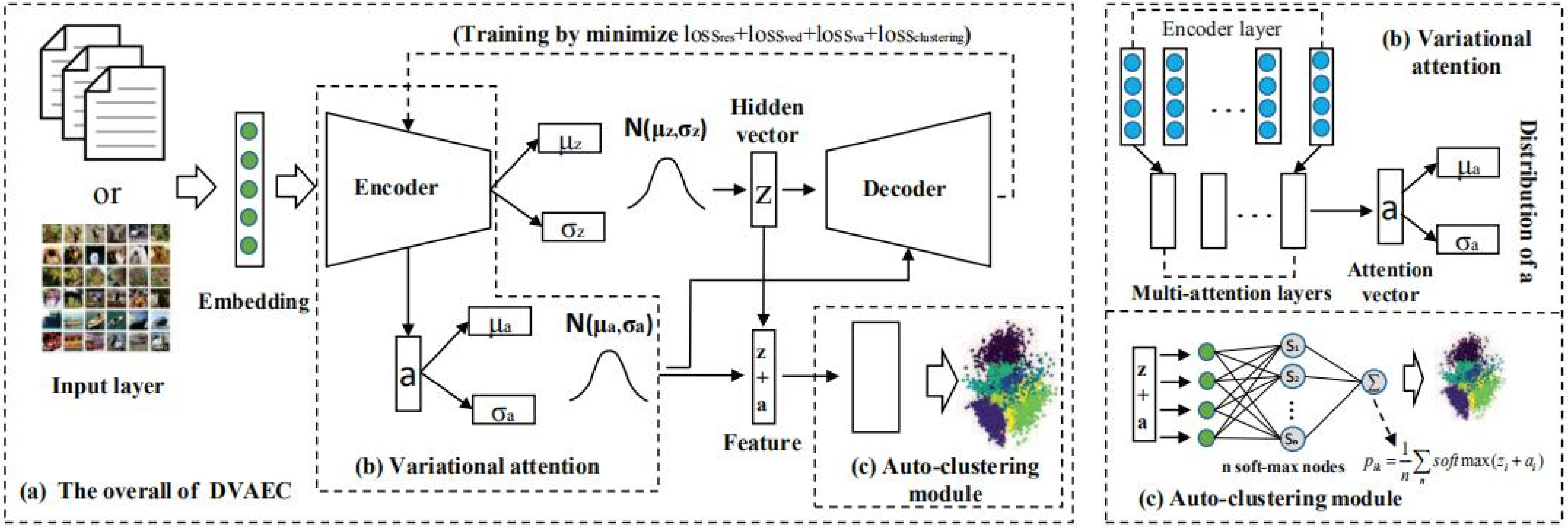
Our framework



- **Input layer**: It is to reshape the data as a input of the model.
- **Encoder layer**: Encoding layer consists of multi-convolutional layers(CNN)
- **Calculation layer**: It calculates the mean and variance of the encoder’s feature.
- **Variational attention layer**: This layer calculates the attention vector between each embedded encoder layer, and then calculates the mean and variance of the attention vector.

Method

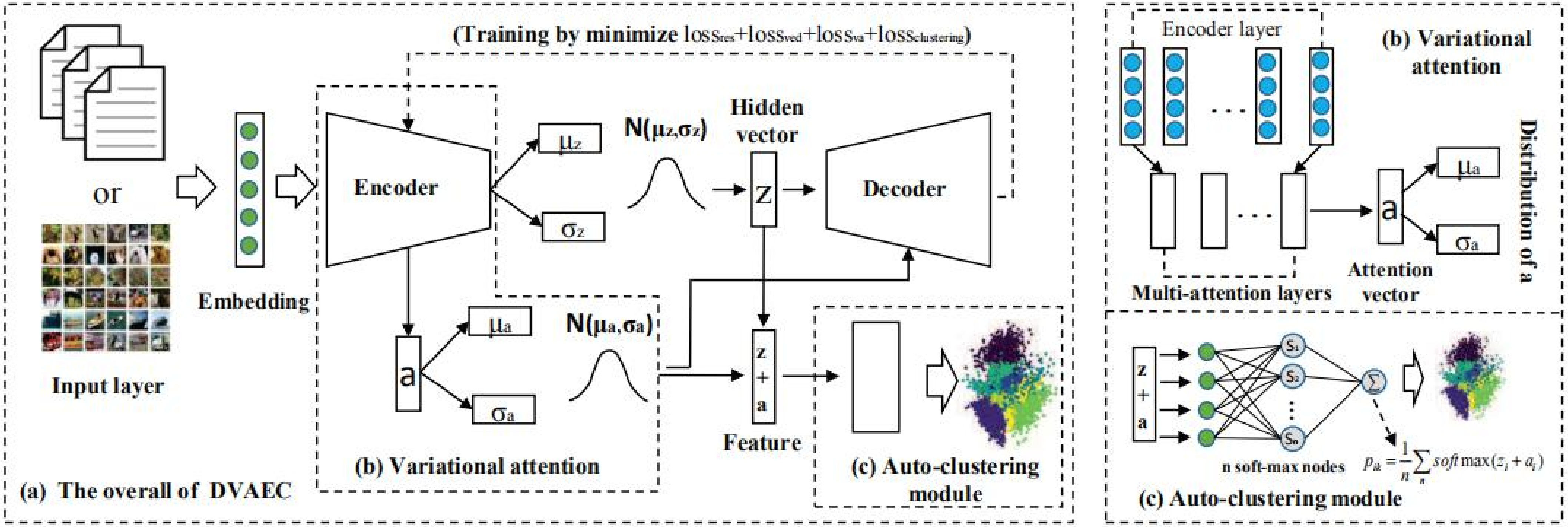
Our framework



- **Hidden layer**: The hidden layer is the hidden space (z) of the model. It connects the calculation layer and decoder layer, which is the feature space and as the input of decoder phase.
- **Decoder layer**: The decoder layer is the corresponding encoder layer, mainly to reconstruct the input according to the deconvolutional networks.
- **Output layer**: The output layer of encoder-decoder framework is to reconstruct the input of the encoder.

Method

Our framework



- **Feature-aware auto-clustering module**: We design a feature-aware auto-clustering module to learn similarity calculation and predict category directly instead of traditional sample similarity calculation process.

Contents

Experiments

- To verify the effectiveness of the our DVAEC, we introduce the experiments settings, and then analyze the experimental results compared with several popular methods.

Dataset	N	k	size	V.
Fashion-MNIST	60k	10	$28\times28\times1$	-
CIFAR-10	60k	10	$28\times28\times3$	-
USPS	9.3k	10	$16\times16\times1$	-
20news	3k	4	-	2.8k
REUTERS	10k	10	-	13k
StackOverflow	20k	20	-	23k

K-means based comparison methods:

K-means+AE、K-means+VAE、K-means+Our(z)、K-means+Our(z+a).

Popular comparison methods (deep clustering methods):

DEC、IDEC、VaDE、SpectralNet.

Experiments

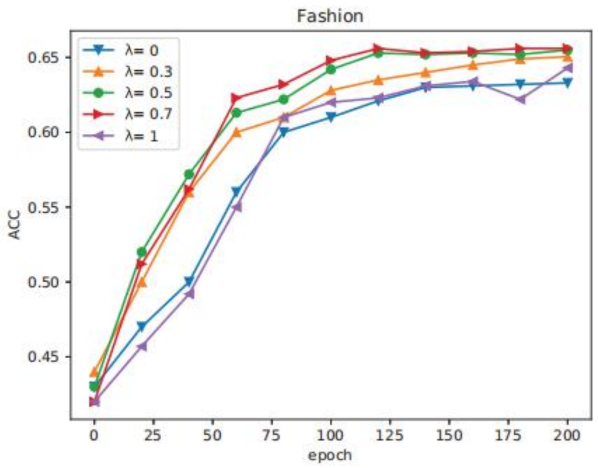
- To evaluate the performance of DVAEC, we compare it with several popular baseline methods. we set three comparison methods based on K-means. For the accuracy of the overall clustering task, we select 4 popular comparison models.

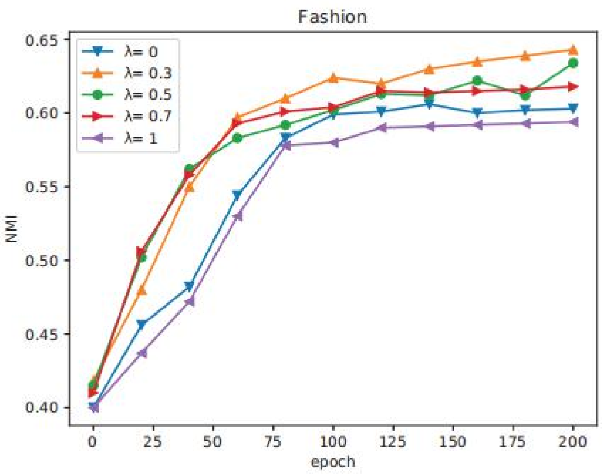
Datasets / Methods	Fashion		CIFAR-10		USPS		20news		REUTERS		StackOverflow	
	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
K-means+AE	47.5	51.5	21.8	10.1	65.3	62.8	33.76	0.71	53.3	52.7	40.1	39.2
K-means+VAE	50.3	51.4	28.0	23.4	70.3	61.1	45.3	23.1	73.5	70.7	42.7	40.5
K-means+our(z)	52.0	51.6	31.2	32.6	71.8	68.4	51.8	47.1	77.2	74.3	45.3	44,7
K-means+our($z+a$)	53.4	53.6	32.3	31.7	73.3	72.5	51.9	49.0	78.6	76.3	46.1	45.2
DEC [4]	51.6	54.6	26.3	25.7	74.1	74.3	50.1	44.4	75.6	70.4	46.3	45.6
IDEC [10]	52.9	55.7	25.1	24.7	76.2	78.5	53.6	44.5	-	-	47.9	46.2
VaDE [8]	58.2	57.3	44.6	45.2	76.9	71.2	67.4	63.5	79.2	72.7	47.1	46.5
SpectralNet [21]	60.1	58.7	48.3	**46.7**	82.5	**80.4**	73.4	71.5	82.1	80.0	52.7	50.4
DVAEC (ours)	**64.7**	**61.4**	**50.6**	42.6	**84.5**	79.3	**75.1**	**70.0**	**83.5**	**81.7**	**56.4**	**53.7**

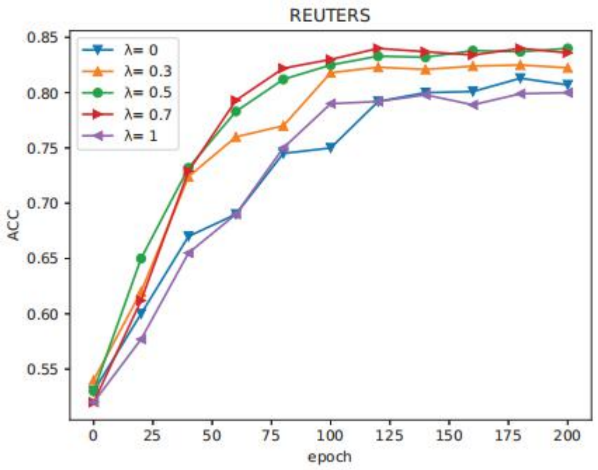
Experiments

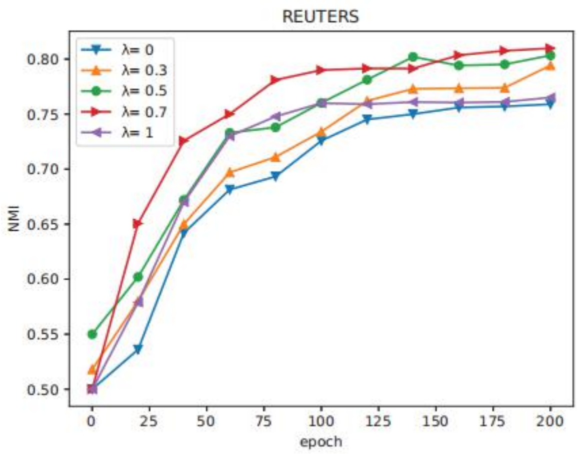
Parameter sensitivity analysis

Impact of the parameter λ on the REUTERS dataset (text). It shows that the variation of ACC (a) and NMI (b) with epoch sizes. The main contrast is the influence of parameter λ on our model when processing text data. It is limited that only lambda changes, and the remaining parameter settings are consistent.









Contents

Conclusions

In this paper, we aim to incorporate variational attention mechanism to enhance representation learning and complete clustering tasks effectively.

- We introduce variational attention mechanism to enhance the feature representation for deep clustering.
- Besides, we design a novel feature-aware auto-clustering module to predict the cluster of the sample precisely.
- Finally, we propose a joint optimization object for the sake of better jointly optimize feature learning and clustering. Extensive experiments on six real datasets demonstrate that our method outperforms several popular models.

Thanks