USAD : Unsupervised Anomaly Detection on Multivariate Time Series

Julien Audibert
julien.audibert@orange.com
EURECOM
Biot, France
Orange
Sophia Antipolis, France

Pietro Michiardi
pietro.michiardi@eurecom.fr
EURECOM
Biot, France

Frédéric Guyard
frederic.guyard@orange.com
Orange Labs
Sophia Antipolis, France

Sébastien Marti
sebastien.marti@orange.com
Orange
Sophia Antipolis, France

Maria A. Zuluaga
zuluaga@eurecom.fr
EURECOM
Biot, France

KDD 2020

Total cites :8

2021.06.09

최영제

Index

Introduction

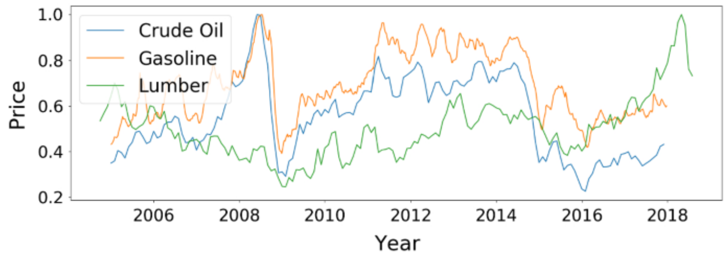
Unsupervised anomaly detection on multivariate time series

기존 관측과는 상이하여 다른 매커니즘에 의해 생성되었다고 판단할만한 관측값

Autoencoder (AE)
Generative adversarial networks (GANs)
…

Unsupervised Anomaly Detection on Multivariate Time Series

각 시간 단위마다 여러 개의 값을 가지는 시계열

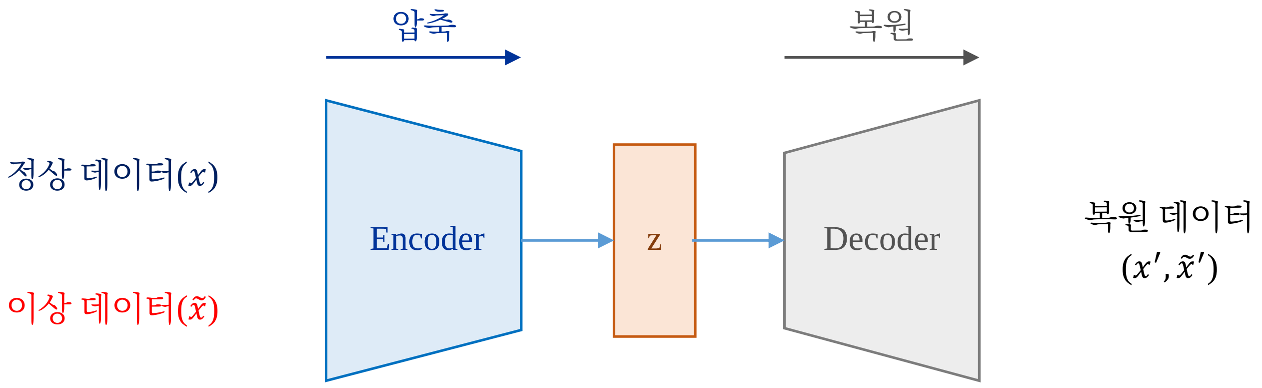


이미지 출처 : https://link.springer.com/article/10.1007/s10994-019-05815-0?shared-article-renderer

Introduction

Unsupervised anomaly detection example

- AE를 활용한 anomaly detection은 데이터를 압축&복원하는 과정에서 발생하는 reconstruction error를 anomaly score로 사용
- 학습에는 정상 데이터만을 사용하기 때문에 이상 데이터가 들어올 경우 큰 reconstruction error가 발생
- 다만 AE는 압축 과정에서 복원에 불필요한 정보를 제거하여 비정상을 탐지할 수 있는 abnormal information이 사라질 가능성이 존재
- 정상 분포와 유사한 비정상 데이터가 들어올 경우 이를 구별할 수 없음(=최대한 정상처럼 복원하기 때문)

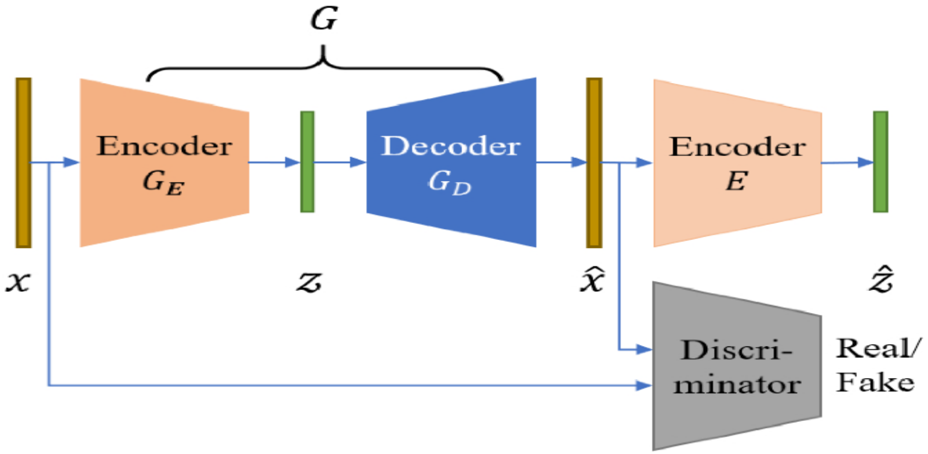


[AE based anomaly detection architecture]

Introduction

Unsupervised anomaly detection example

- GANs based 방법의 경우 fake (abnormal)와 real (normal)을 구분하는 discriminator의 도입으로 정상 데이터만을 활용하더라도 보다 상세하게 비정상을 구분할 수 있음(=abnormal information을 포함)
- Discriminator를 속이기 위해서는 real 정보 뿐 아니라 fake에 대한 정보를 포함하도록 encoder 및 decoder가 학습하기 때문
- 다만 GANs는 안정적인 학습이 어렵다는 단점이 존재함



[GANs based anomaly detection architecture (GANomaly)]

Unsupervised anomaly detection example

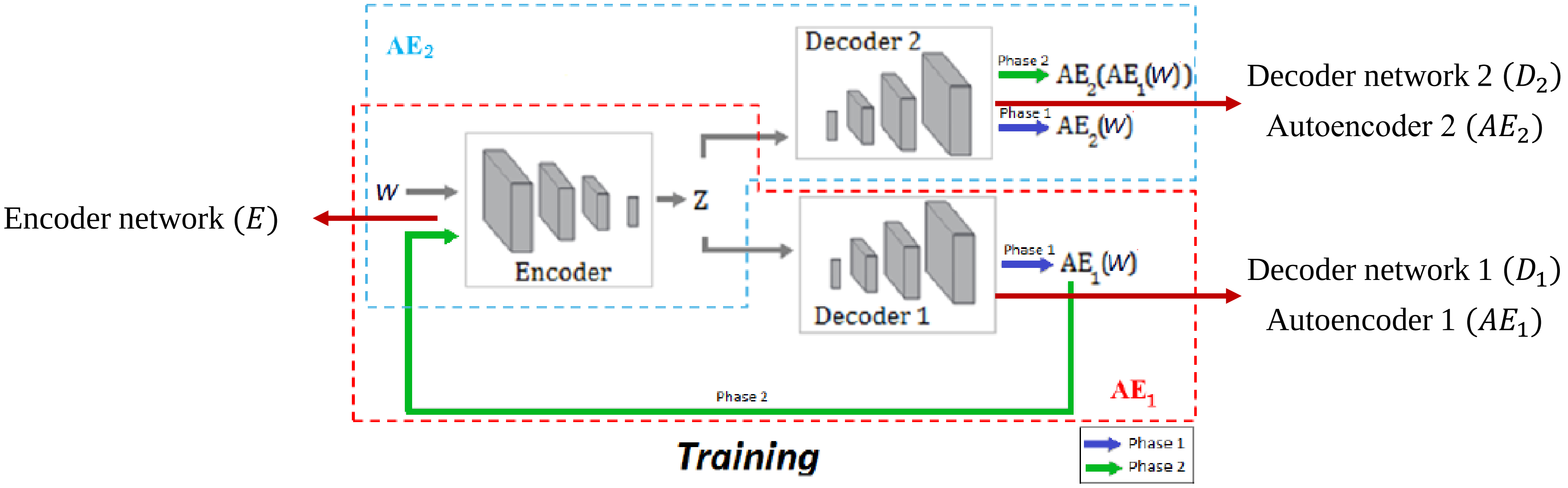
- Unsupervised anomaly detection 알고리즘의 흐름

LSTM-AE → AnoGAN → DAGMM → GANomaly →
mSCRED → RaPP → OmniAnomaly → BeatGAN →
MAD-GAN → f-AnoGAN → adVAE → USAD →
SOM-DAGMM → TadGAN → NVAE-GAN → SIS-VAE

* AE : autoencoder, VAE : variational autoencoder, GANs : generative adversarial networks

USAD

- USAD는 학습이 쉬운 AE의 장점과 abnormal information을 강제할 수 있는 GANs의 장점을 결합한 모델임
- AE를 사용하되 adversarial training을 적용하여 보다 상세한 anomaly detection을 추구하고자 함
- USAD의 architecture는 다음과 같음



[USAD architecture]

Methods

USAD training process

- USAD is trained in two phases, 1) autoenocder training and 2) adversarial training
- 1) Autoencoder training

$$\mathcal{L}_{AE_1} = \|W - AE_1(W)\|_2$$
$$\mathcal{L}_{AE_2} = \|W - AE_2(W)\|_2$$

W : Sequence of windows in training set

- 2) Adversarial training
  - Train AE_2 to distinguish the real data from the data coming from AE_1, and train AE_1 to fool AE_2
  - The objective of AE_1 is to minimize the difference between W and the output of AE_2

$$\min_{AE_1} \|W - AE_2(AE_1(W))\|_2$$

  - The objective of AE_2 is to maximize this difference

$$\max_{AE_2} \|W - AE_2(AE_1(W))\|_2$$

2 Methods

USAD training process

- 각각의 AE의 최종 loss는 다음과 같음

Reconstruction error on real data

AE_2's reconstruction error on fake data

$$\mathcal{L}_{AE_1} = \underline{\frac{1}{n}\|W - AE_1(W)\|_2} + \left(1 - \frac{1}{n}\right) \underline{\|W - AE_2(AE_1(W))\|_2}$$

$$\mathcal{L}_{AE_2} = \underline{\frac{1}{n}\|W - AE_2(W)\|_2} - \left(1 - \frac{1}{n}\right) \underline{\|W - AE_2(AE_1(W))\|_2}$$

n : Training epochs

$AE_1(W)$: Fake (abnormal)

- AE_1의 경우 real에 대한 reconstruction error와 fake에 대한 AE_2의 reconstruction error 둘 모두 최소일 때 최소
- AE_2의 경우 real에 대한 reconstruction error가 최소이고 fake에 대한 AE_2의 reconstruction error가 최대일 때 최소
- 정리하면 AE_1는 AE_2가 fake와 real을 구분하지 못하게 만들고, AE_2는 fake이 들어왔을 때 reconstruction error를 크게 만들도록 학습
- 즉 AE_2는 정상 데이터와 비정상 데이터의 미세한 차이를 극대화 시키는 역할을 함

Methods

USAD detection process

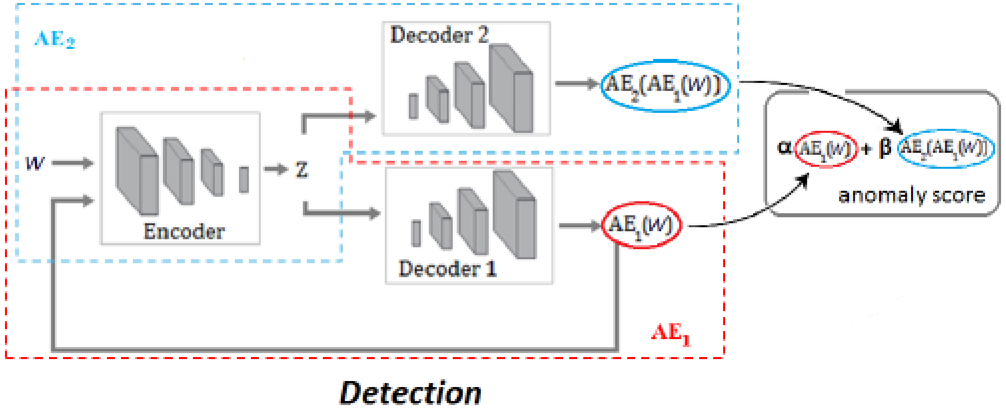
- 위와 같이 학습된 USAD의 anomaly score 산출식은 다음과 같음

$$\mathcal{A}(\widehat{W}) = \alpha\|\widehat{W} - AE_1(\widehat{W})\|_2 + \beta\|\widehat{W} - AE_2(AE_1(\widehat{W}))\|_2$$

$\widehat{W}$: Sequence of windows in test set

$\alpha + \beta = 1$

- 정상과 매우 유사한 비정상 데이터가 들어왔을 때도 USAD는 이를 탐지할 수 있음



[USAD detection process]

Parameter setting	# of detection	Detection sensitivity
$\alpha > \beta$	Reduce	Low
$\alpha < \beta$	Increase	High

[Variation by parameter setting]

Experiments and Results

Experimental setup

- 5 public datasets as follow
  - Secure water treatment (SWatT) : https://github.com/JulienAu/Anomaly_Detection_Tuto
  - Water distribution (WADI) : https://itrust.sutd.edu.sg/itrust-labs_datasets/dataset_info/#wadi
  - Server machine dataset (SMD) : https://github.com/smallcowbaby/OmniAnomaly
  - Soil moisture active passive (SMAP)
  - Mars science laboratory (MSL)

- Evaluation metrics as follow

$$P = \frac{TP}{TP + FP} \qquad R = \frac{TP}{TP + FN} \qquad F1 = \frac{P * R}{P + R} \qquad F1^* = 2\frac{\bar{P} * \bar{R}}{\bar{P} + \bar{R}}$$

$\bar{P}, \bar{R}$: Average precision and recall

- F1 score가 가장 높도록 thresholding을 진행

Experiments and Results

Results

- Point-adjust : detect each observation/time-point independently and assigns a label to single time-point (without/with)

Methods	SWaT						WADI					
	Without			With			Without			With		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
AE	0.9903	0.6295	0.7697	0.9913	0.7040	0.8233	0.9947	0.1310	0.2315	0.3970	0.3220	0.3556
IF	0.9512	0.5884	0.7271	0.9620	0.7315	0.8311	0.2992	0.1583	0.2071	0.6241	0.6155	**0.6198**
LSTM-VAE	0.9897	0.6377	0.7756	0.7123	0.9258	0.8051	0.9947	0.1282	0.2271	0.4632	0.3220	0.3799
DAGMM	0.4695	0.6659	0.5507	0.8292	0.7674	0.7971	0.0651	0.9131	0.1216	0.2228	0.1976	0.2094
OmniAnomaly	0.9825	0.6497	0.7822	0.7223	0.9832	0.8328	0.9947	0.1298	0.2296	0.2652	0.9799	0.4174
USAD	0.9851	0.6618	**0.7917**	0.9870	0.7402	**0.8460**	0.9947	0.1318	**0.2328**	0.6451	0.3220	0.4296

Methods	SMD				SMAP				MSL			
	P	R	F1	F1*	P	R	F1	F1*	P	R	F1	F1*
AE	0.8825	0.8037	0.8280	0.8413	0.7216	0.9795	0.7776	0.8310	0.8535	0.9748	0.8792	0.9101
IF	0.5938	0.8532	0.5866	0.7003	0.4423	0.5105	0.4671	0.4739	0.5681	0.6740	0.5984	0.6166
LSTM-VAE	0.8698	0.7879	0.8083	0.8268	0.7164	0.9875	0.7555	0.8304	0.8599	0.9756	0.8537	0.9141
DAGMM	0.6730	0.8450	0.7231	0.7493	0.6334	0.9984	0.7124	0.7751	0.7562	0.9803	0.8112	0.8537
OmniAnomaly	0.9809	0.9438	**0.9441**	**0.9620**	0.7585	0.9756	0.8054	0.8535	0.9140	0.8891	0.8952	0.9014
USAD	0.9314	0.9617	0.9382	0.9463	0.7697	0.9831	**0.8186**	**0.8634**	0.8810	0.9786	**0.9109**	**0.9272**

Experiments and Results

Results

- 알고리즘별 전체 datasets의 평균 성능(± standard deviation)은 다음과 같음

	P	R	F1	F1*
AE	0.77(0.21)	0.76(0.24)	0.73(0.19)	0.86 (0.04)
IF	0.64(0.17)	0.68(0.11)	0.62(0.12)	0.60 (0.09)
LSTM-VAE	0.72(0.15)	0.80(0.25)	0.75 (0.18)	0.86 (0.04)
DAGMM	0.62(0.21)	0.76(0.29)	0.65(0.22)	0.79 (0.04)
OA	0.73(0.25)	**0.95(0.04)**	0.78(0.19)	0.91(0.04)
USAD	**0.84(0.12)**	0.80(0.25)	**0.79(0.18)**	**0.91(0.04)**

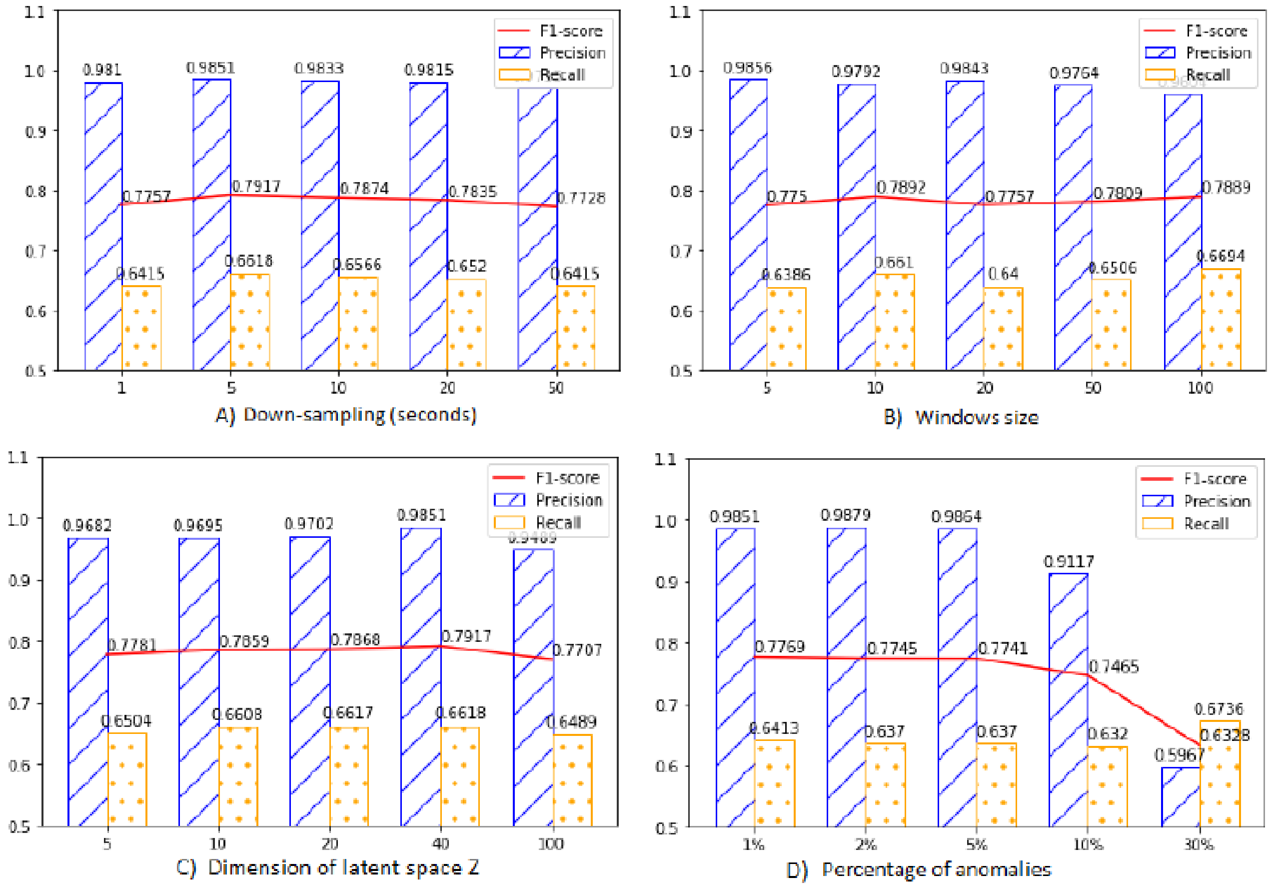
- α와 β에 따른 SWaT dataset의 성능은 다음과 같음

α	β	FP	TP	F1
0.0	1.0	604	35,616	0.7875
0.1	0.9	580	35,529	0.7853
0.2	0.8	571	35,285	0.7833
0.5	0.5	548	34,590	0.7741
0.7	0.3	506	34,548	0.7738
0.9	0.1	299	34,028	0.7684

Experiments and Results

Results – ablation study

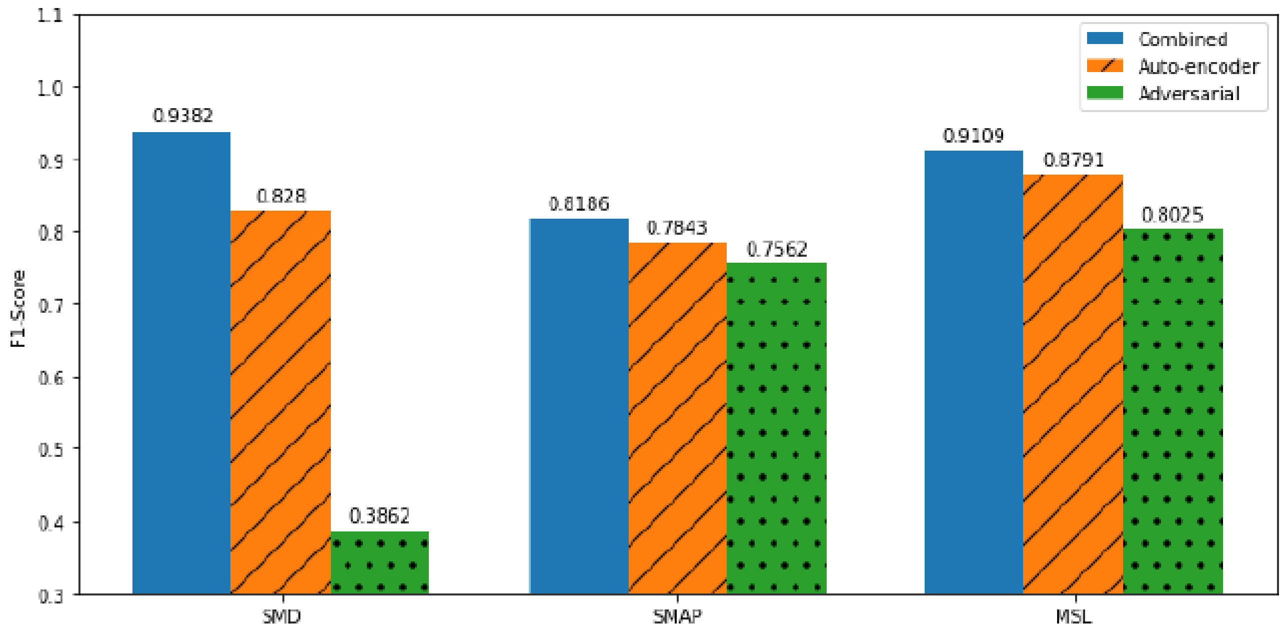
- Down-sampling / windows size / latent space's dimension / percentage of anomalies에 따른 ablation study의 결과



Experiments and Results

Results – ablation study

- With/without adversarial training



Q&A