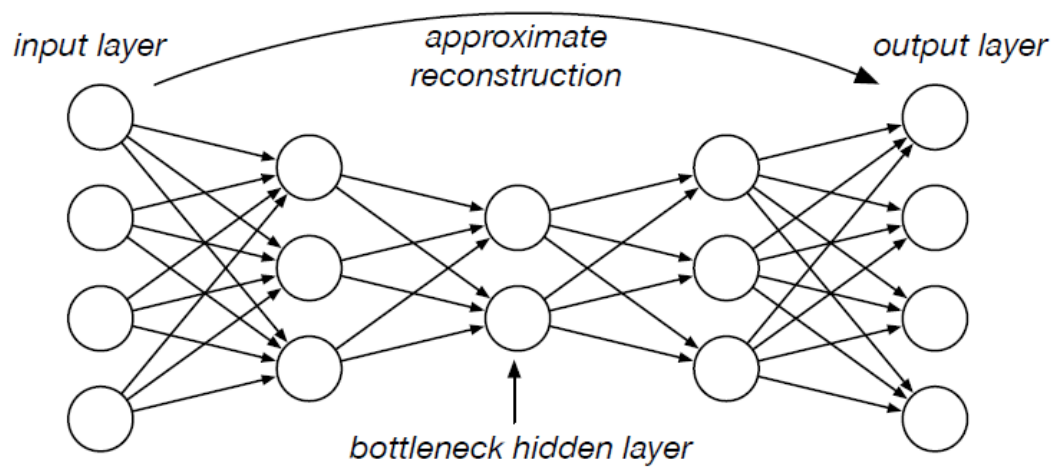


Outlier Detection with Autoencoder Ensembles (2017)

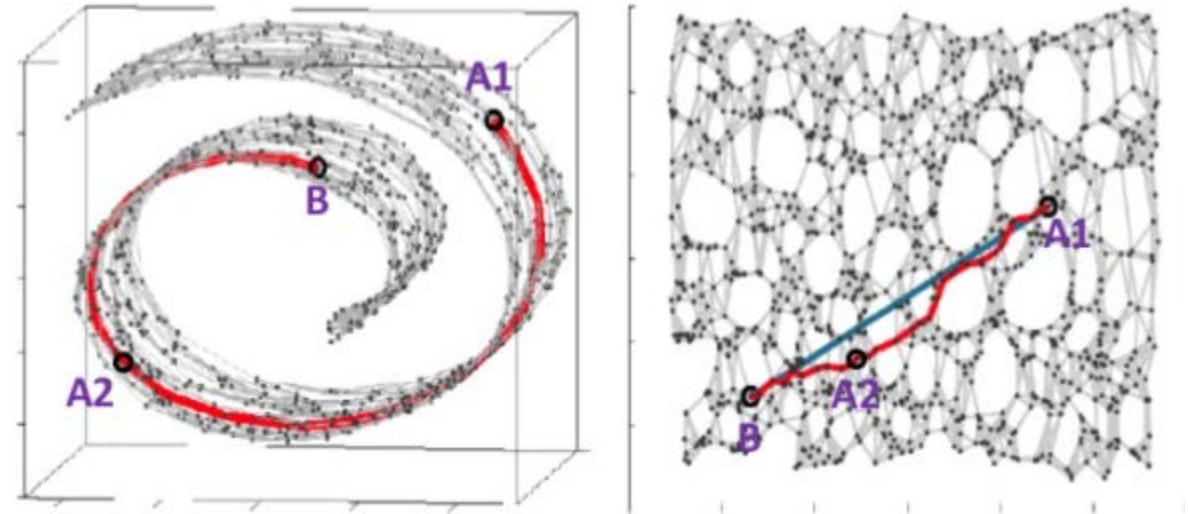
Jinghui Chen, Saket Sathey, Charu Aggarwaly, Deepak Turagay

AutoEncoder

- Unsupervised learning
- Nonlinear dimensionality reduction
- Feature extraction
- Manifold learning



- $\min (x_i - \hat{x}_i)^2$
- input=output



<https://www.slideshare.net/NaverEngineering/ss-96581209>

AutoEncoder in Anomaly detection

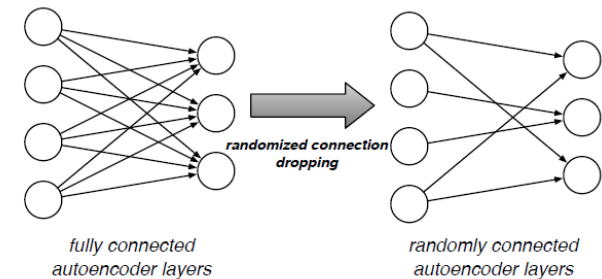
- Learns the representations of data instances by optimizing a general feature learning objective function that is not primarily designed for anomaly detection but empower the anomaly detection
- 장점 : 다양한 data에 사용가능
- 단점 : feature representation의 불완전성, overfitting

Introduction

- Autoencoder encoder ensembles for unsupervised outlier detection
 - 저 자 들 은 RandNet (Randomized Neural Network for Outlier Detection) 이라고 이름을 지었다.

RandNet Model

- Basic idea
 - Outlier들은 AE를 통해 reconstruction할 때 더 큰 error를 가질 것
- Outlier들이 training set에 함께 있고 NN의 고질적인 문제인 overfitting이 일어날 수 있다.
 - 이를 위해 Ensemble방법을 사용하였다.
 - 여러 개의 AE model을 만들고 다양성을 위해 각 model은 random 하게 connection이 drop된다.



NN Structure

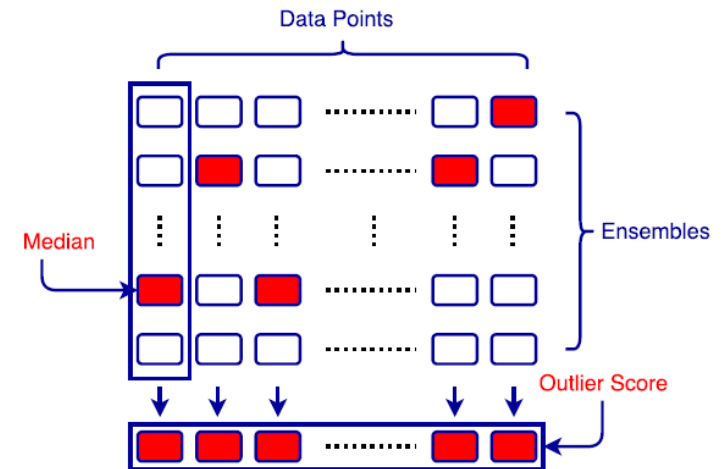
- AE에서 점점 node의 수가 줄어드는 형태인데 이를 정하는 parameter를 'structure parameter'라 부르고 이전 node의 수에 비해 해당 비율만큼 줄였다.
- Choice of Activation Functions
 - Sigmoid, ReLU
 - 'Dying ReLU' 문제때문에 마지막 두개의 layer에서는 sigmoid사용
 - Gradient가 너무 클 때, 주로 문제가 발생하는데 sigmoid의 gradient vanishing으로 완충

NN Structure

- Random Connection Generation
 - 인접한 layer가 각 a, b 개의 node를 갖고 있는 경우 connection들은 총 ab 개가 된다.
 - ab 개의 connection들을 choose with replacement
 - 이를 통해 뽑힌 connection만 살린다.
 - So, can produce different types of randomly connected autoencoders with varying densities

Outlier Scoring

- Outlier score는 AE의 reconstruction error로 구한다.
 - Data point마다 score를 구할 수 있다.
- 각 AE model마다 Outlier score를 구한 뒤에 median 값을 최종 Outlier score로 선택한다.
- 각 model마다 score에서 차이가 존재할 수 있기에 각각 normalize하여 동등한 비교가 가능하게 한다.
- $[OS_i]_j = \sum_{k=1}^d \left([x_{ij}]_k - [\hat{x}_{ij}]_k \right)^2$
 - $i = 1, \dots, m$ $j = 1, \dots, n$ (m개 model, n개 data)
 - $x_{ij} \in R^d$

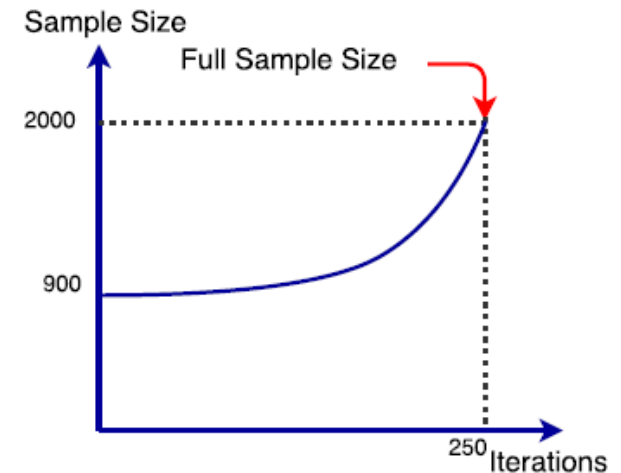


Training and Adaptive Sampling

- Adaptive Learning Approach
 - RMSprop를 사용
- Pre-training
 - 3-layer NN model으로 학습
 - 여기서 구해진 weight를 initial weight로 사용하여 더 깊은 network를 효율적으로 학습

Training and Adaptive Sampling

- Adaptive Sampling
 - Make the optimization procedure more efficient by changing (adapting) the training sample size in each iteration.
 - 학습초반에는 정교한 gradient가 필요하기보다 대략적으로 방향성만 맞으면 된다. 진행되면서 점차 정확한 gradient가 필요하다.
 - (근데 최근까지도 batch size가 작은게 좋은지 크게 좋은지 완벽하게 밝혀진 연구는 없다고 한다.)



Experimental Results, Conclusions

- Ensemble components들은 25개부터 200개까지 실험해봤는데 많을수록 좋아지는 편이지만 그 정도가 크지 않다.
- Layer의 수는 3개부터 9개까지 실험했는데 깊은 model이 성능이 더 좋았다. 다만 깊은 model은 훈련이 어렵고 data의 수가 적은 경우 성능이 오히려 떨어지는 경우도 있었다. 저자는 7개 layer를 추천한다고 한다.
- Overfitting을 ensemble approach로 완화시켰다.

마무리

- AE가 anomaly detection분야에서 어떤 방식으로 응용되는지 알 수 있었다.
- AE의 성능은 적절한 node의 수와 적절한 layer의 수가 중요하다.
- Theoretical한 부분은 없지만 Ensemble기법을 사용한 아이디어는 재밌었다.
- 예전에 비해 많은 부분이 해결되었지만 deep한 model의 훈련의 어려움과 overfitting 문제는 딥러닝을 공부하면서 항상 유의하자.