Timeseries Anomaly Detection using Temporal Hierarchical One-Class Network

한양대학교 산업공학과
IDSL 석사과정 고경준

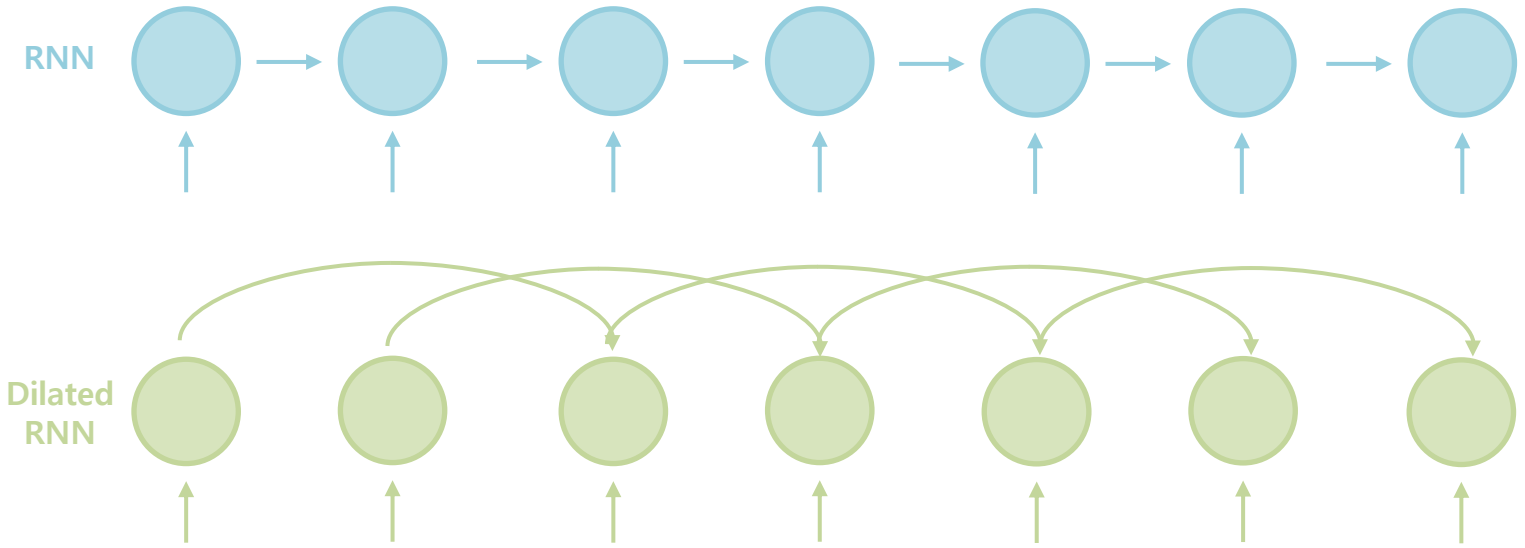
HANYANG UNIVERSITY

Contents

1. Dilated-RNN
2. Deep SVDD
3. THOC (Temporal Hierarchical One-Class Network)
4. Experiments
5. Ablation Study

1. Dilated RNN

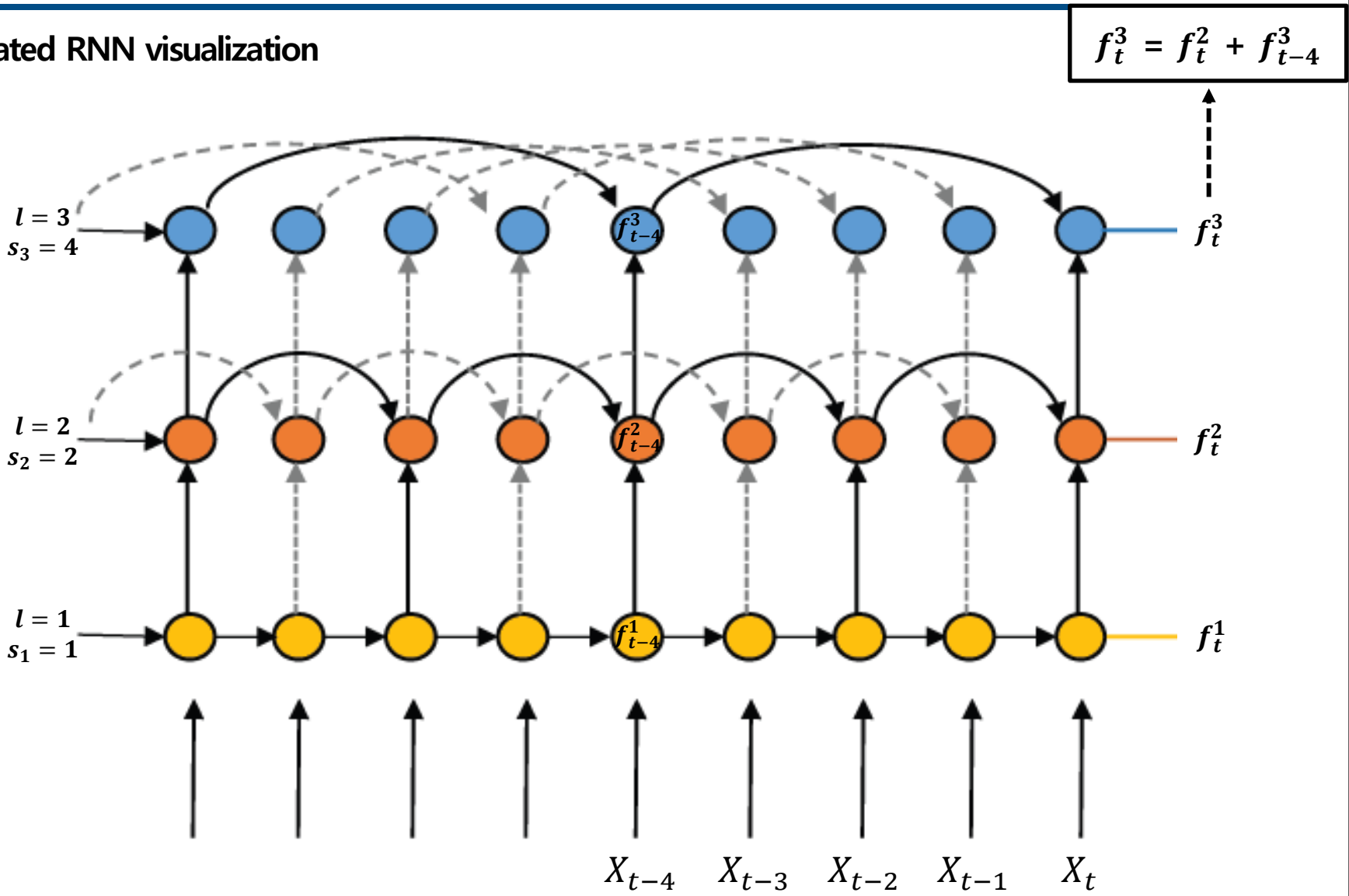
Dilated RNN



- 바로 이전 시점의 정보를 반영하지 않고, skip length(s^l) 시점 이전의 정보를 반영하는 skip connection 개념을 도입
- Layer의 층 수(l)에 따라 skip length를 지수적으로 변경($s^l = 2^{l-1}$)
- 이러한 Layer를 여러 겹 쌓아서 short / long term dependency를 반영

1. Dilated RNN

Dilated RNN visualization



2. Deep SVDD

Deep SVDD

- 딥러닝을 기반으로 재표현된 feature space에서 정상 데이터를 둘러싸는 가장 작은 구를 찾고, 해당 경계면을 기반으로 이상치를 탐지

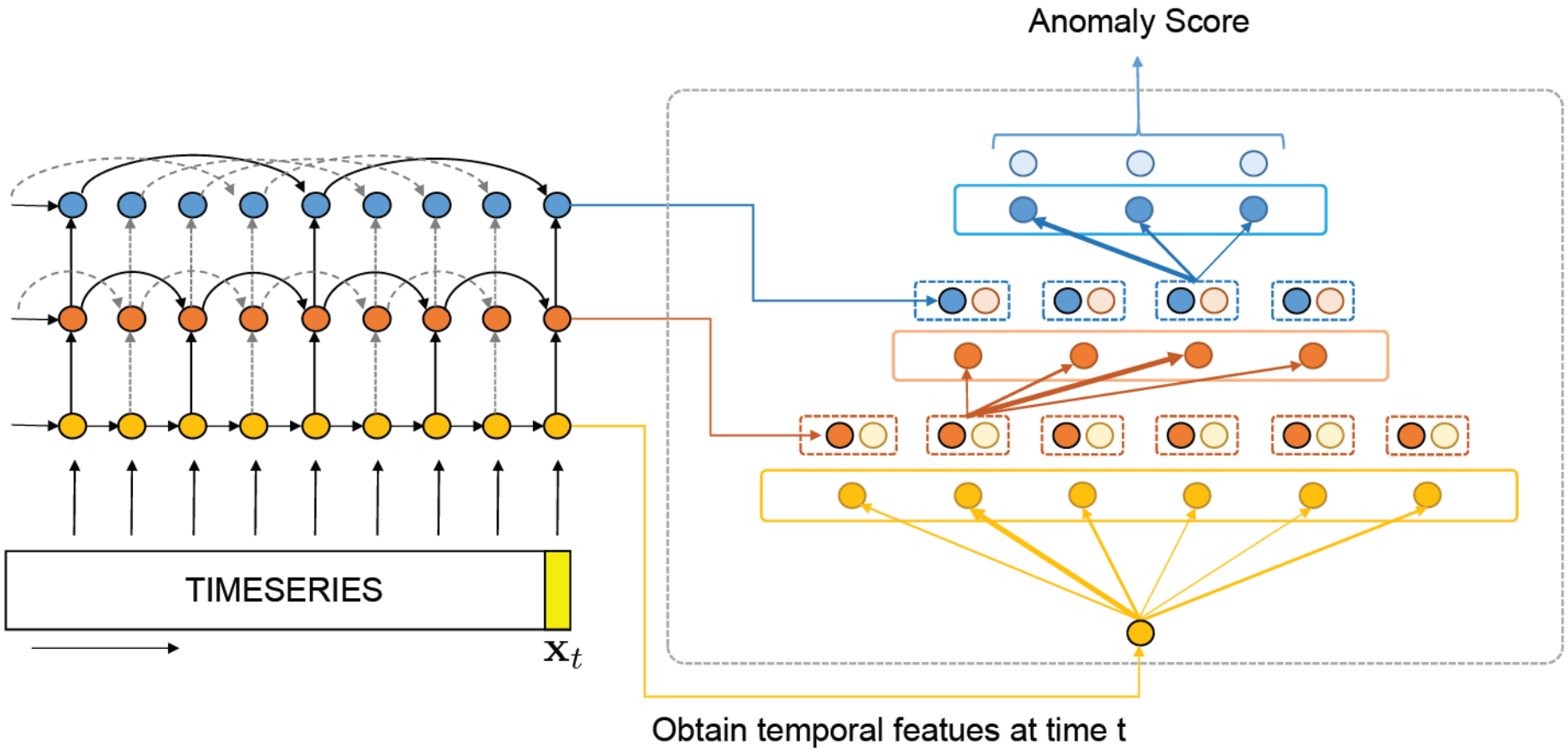
$$\min_{R,\mathcal{W}} R^2 + \frac{1}{\nu N}\sum_{i=1}^{N}\max\{0, \|\mathrm{NN}(\mathbf{x}_i;\mathcal{W}) - \mathbf{c}\|^2 - R^2\} + \lambda\Omega(\mathcal{W}),$$

첫번째 term : 구의 부피를 최소화
두번째 term : 구의 바깥에 찍힌 data point에 penalty
세번째 term : Regularization

$$= \min_{\mathcal{W}} \frac{1}{N}\sum_{i=1}^{N}\|\mathrm{NN}(\mathbf{x}_i;\mathcal{W}) - \mathbf{c}\|^2 + \lambda\Omega(\mathcal{W}).$$

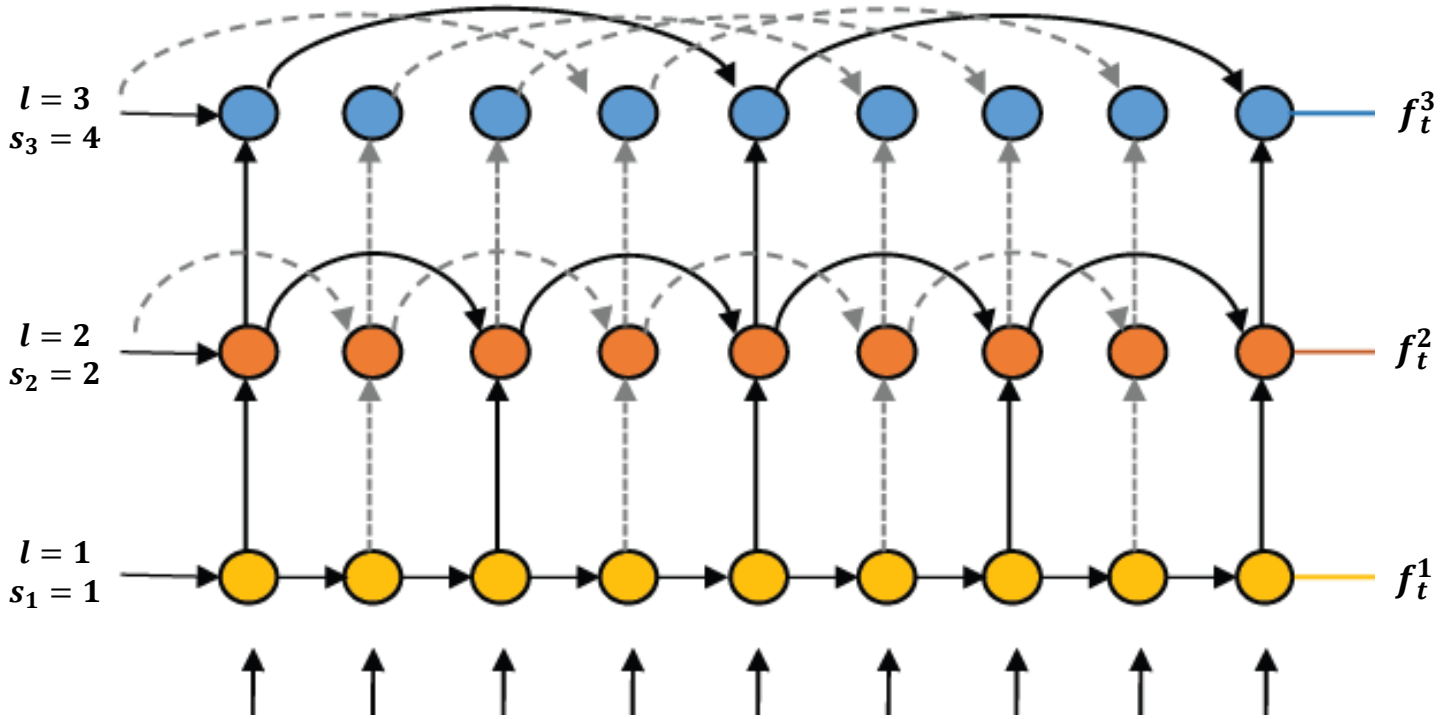
3. THOC

Architecture



3. THOC

3.1.1 Multiscale Temporal Features



$$\mathbf{f}_t^l = \begin{cases} \mathcal{F}_{\text{RNN}}(\mathbf{x}_t, \mathbf{f}_{t-s^{(l)}}^l) & \text{if } l = 1, \\ \mathcal{F}_{\text{RNN}}(\mathbf{f}_t^{l-1}, \mathbf{f}_{t-s^{(l)}}^l) & \text{otherwise,} \end{cases} \qquad s^{(l)} = \bar{M}_0 \prod_{i=1}^{l-1} \bar{M} \text{ (in Figure 1, } \bar{M}_0 = 1 \text{ and } \bar{M} = 2).$$

3. THOC

3.1.2 Fusing the Multiscale Features

<Step 1. Assignment>

- Hierarchical clustering procedure에는 layer(l)별로 K^l개의 cluster가 존재
- 이전 layer의 output($\bar{f}_{t,i}^{l-1}$)을 input으로 받아서 cluster에 배정될 확률($P_{t,i\to j}^{l}$)을 계산

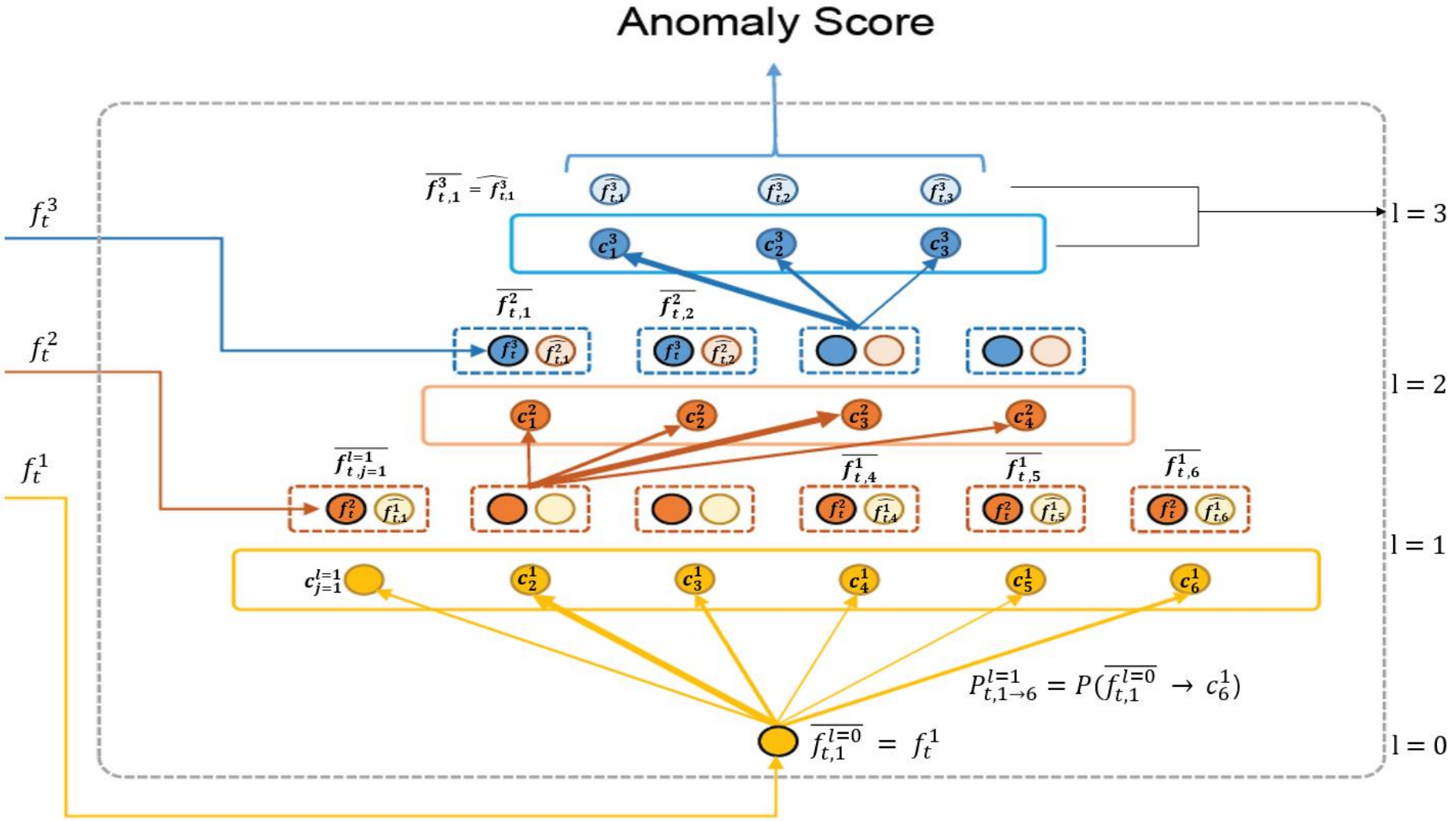
$$P_{t,i\to j}^{l} = P(\bar{\mathbf{f}}_{t,i}^{l-1} \to \mathbf{c}_j^l) = \frac{\exp(\text{score}(\bar{\mathbf{f}}_{t,i}^{l-1}, \mathbf{c}_j^l)/\tau)}{\sum_{k=1}^{K^l} \exp(\text{score}(\bar{\mathbf{f}}_{t,i}^{l-1}, \mathbf{c}_k^l)/\tau)}, \quad (5)$$

(τ : temperature parameter)

$$\text{score}(\bar{\mathbf{f}}, \mathbf{c}) = \mathbf{f}^\top \mathbf{c}/(\|\mathbf{f}\| \cdot \|\mathbf{c}\|). \quad (6)$$

3. THOC

3.1.2 Fusing the Multiscale Features



3. THOC

3.1.2 Fusing the Multiscale Features

<Step 2. Update>

- Cluster에 배정될 확률을 이용하여 다음 layer에 전달 될 output이 계산됨
- assignment와 update가 반복되면 마지막 layer(L)에서 output($\bar{f}_{t,i}^{L}$)이 계산됨

$$\hat{\mathbf{f}}_{t,j}^{l} = \sum_{i=1}^{K^{l-1}} P_{t,i\to j}^{l} \mathrm{ReLU}(\mathbf{W}^{l}\bar{\mathbf{f}}_{t,i}^{l-1} + \mathbf{b}^{l}), \quad j = 1, \ldots, K^{l}, \tag{7}$$

$$\bar{\mathbf{f}}_{t,j}^{l} = \begin{cases} \mathbf{f}_{t}^{1} & \text{if } l = 0 \\ \mathcal{F}_{\mathtt{MLP}}([\hat{\mathbf{f}}_{t,j}^{l}; \mathbf{f}_{t}^{l+1}]) & \text{if } 1 \le l \le L-1 \\ \hat{\mathbf{f}}_{t,j}^{L} & \text{otherwise (i.e., } l = L) \end{cases}. \tag{8}$$

3. THOC

3.2 Multiscale Support Vector Data Description (MVDD)

- 세 가지 Loss의 합으로 구성됨
- 첫번째 Loss (L_{THOC}) : 할당 확률에 가중치를 준 클러스터 center와의 거리

$$\mathcal{L}_{\text{THOC}} = \frac{1}{N K^L} \sum_{s=1}^{N} \frac{1}{T_s} \sum_{t=1}^{T_s} \sum_{j=1}^{K^L} R_{t,j,s}^{L} d(\bar{\mathbf{f}}_{t,j,s}^{L}, \mathbf{c}_j^L) + \lambda \Omega(\mathcal{W}), \tag{9}$$

N : 시계열 변수 X_s (X_1 , … , X_N)의 개수
T_s : 시계열 변수 X_s (X_1 , … , X_N)의 길이
K^L : hierarchical procedure의 마지막 layer(L)의 클러스터 개수

$$R_{t,j}^{l} = \frac{\exp(\tilde{R}_{t,j}^{l})}{\sum_{i=1}^{K^l} \exp(\tilde{R}_{t,i}^{l})}, \text{ where } \tilde{R}_{t,j}^{l} = \begin{cases} P_{t,i \to j}^{1}, & \text{if } l = 1 \\ \sum_{i=1}^{K^{l-1}} P_{t,i \to j}^{l} R_{t,i}^{l-1} & \text{if } 1 < l \leq L \end{cases}. \tag{10}$$

3. THOC

3.2 Multiscale Support Vector Data Description (MVDD)

- 두번째 Loss (L_{orth}) : cluster center의 다양성을 확보하기 위해 orthogonal하게 만듦

$$\mathcal{L}_{\text{orth}} = \frac{1}{L}\sum_{l=1}^{L} \|(\mathbf{C}^l)^\top \mathbf{C}^l - \mathbf{I}\|_F^2, \tag{11}$$

- 세번째 Loss (L_{TSS}) : 

$$\mathcal{L}_{\text{TSS}} = \frac{1}{NL}\sum_{s=1}^{N}\sum_{l=1}^{L}\left(\frac{1}{T_s - s^{(l)}}\sum_{t=s^{(l)}+1}^{T_s} \|\mathbf{W}_{\text{pred}}^l \mathbf{f}_{t-s^{(l)},s}^l - \mathbf{x}_{t,s}\|^2\right). \tag{12}$$

3. THOC

3.2 Multiscale Support Vector Data Description (MVDD)

- **Total loss**

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{THOC}} + \lambda_{\text{orth}}\mathcal{L}_{\text{orth}} + \lambda_{\text{TSS}}\mathcal{L}_{\text{TSS}}, \tag{13}$$

- **AnomalyScore(x_t)**

$$\text{AnomalyScore}(\mathbf{x}_t) = \sum_{j=1}^{K^L} R_{t,j}^L \cdot d(\bar{\mathbf{f}}_t^L, \mathbf{c}_j^L).$$

(predefined threshold δ와 비교되어 abnormal 여부를 판별)

4. Experiments

4.1 Data Sets

	dim	length	#training	#validation	#testing
2D-gesture	2	80	8,170	420	2,420
power-demand	1	80	18,145	4,786	10,000
KDD-Cup99	34	100	56,139	24,601	24,602
SWaT	51	100	47,420	22,396	22,396
MSL	55	100	40,721	17,396	73,629
SMAP	25	100	94,528	40,455	427,517

(i) *2D-gesture* [9], which records the X-Y coordinate sequences of hand gestures in a video;

(ii) *Power demand*[9], which contains a year of power demand at a Dutch research facility;

(iii) *KDD-Cup99* data from the DARPA'98 Intrusion Detection Evaluation Program [14]. It contains around seven million network traffic connection records over a 7-week period. A connection is a sequence of TCP packets. Each record is labeled as either normal or attack;

(iv) Secure Water Treatment (*SWaT*) data [18], which is collected from a water treatment testbed over 11 days. 36 attacks were launched during the last 4 days of the collection process. These attacks were launched with different intents and diverse lasting durations (from a few minutes to an hour);[1]

(v) Mars Science Laboratory rover (*MSL*); and

(vi) Soil Moisture Active Passive satellite (*SMAP*) data: Both *MSL* and *SMAP* are public data sets from NASA [8]. They contain telemetry anomaly data derived from the Incident Surprise Anomaly (ISA) reports of spacecraft monitoring systems. Each data set has a training and a testing set. Anomalies in the testing set are labeled, while the training set contains unlabeled anomalies.

4. Experiments

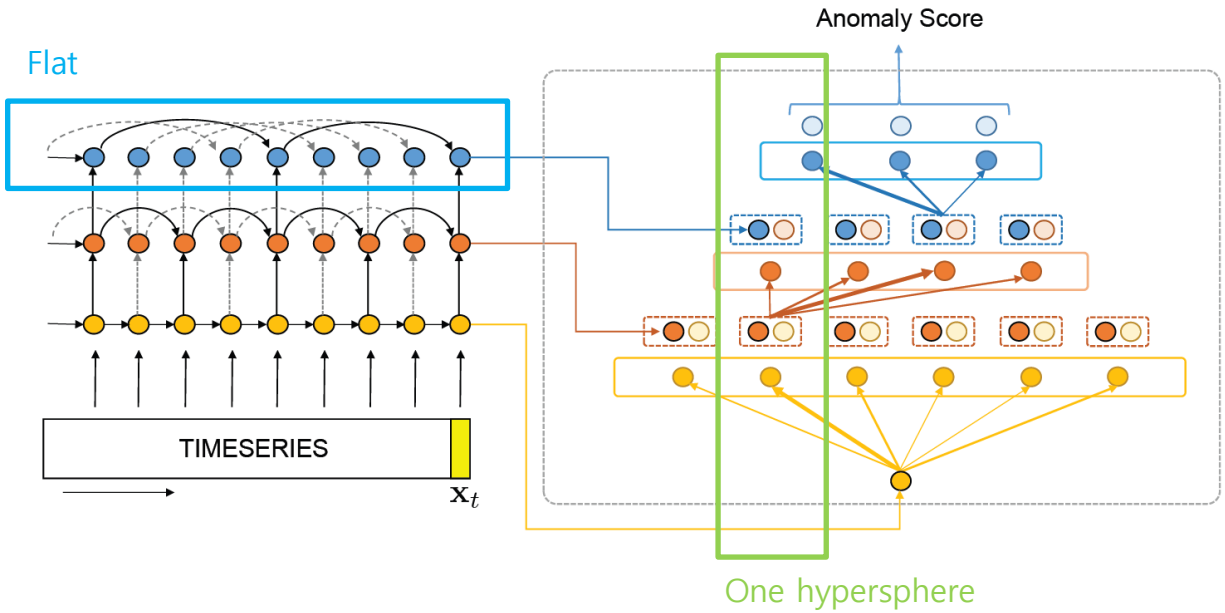
4.2 Baselines for Comparison

		2D-gesture			power-demand			KDD-Cup99			SWaT			avg
		prec	rec	F1	prec	rec	F1	prec	rec	F1	prec	rec	F1	rank
MAD	LOF	27.82	87.21	42.18 (8)	15.29	28.13	19.81 (9)	95.38	99.55	97.42 (11)	76.97	98.36	86.36 (7)	8.75
MAD	OC-SVM	65.50	25.57	36.78 (14)	12.40	60.43	20.58 (8)	95.25	99.92	97.53 (10)	99.47	61.47	75.98 (13)	11.25
MAD	iso forest	28.54	68.04	40.22 (10)	7.85	89.77	14.44 (13)	96.85	99.38	98.10 (7)	99.00	74.47	85.00 (9)	9.75
DAD	deep SVDD	26.26	64.53	37.32 (13)	11.51	64.74	19.54 (10)	89.83	100.0	94.64 (14)	97.68	71.88	82.82 (11)	12
DAD	AnoGAN	57.85	46.50	51.55 (4)	20.28	44.41	28.85 (5)	93.11	99.93	96.40 (12)	99.01	77.01	86.64 (5)	6.5
DAD	DAGMM	25.66	80.47	38.91 (12)	34.37	41.72	37.69 (4)	96.12	99.70	97.86 (8)	90.60	80.72	85.38 (8)	8.0
DAD with timeseries	EncDec-AD	24.88	100.0	39.85 (11)	13.98	54.20	22.22 (6)	89.74	99.50	94.37 (13)	93.69	63.31	75.56 (14)	11
DAD with timeseries	LSTM-VAE	36.62	67.76	47.54 (6)	8.00	56.66	14.03 (14)	98.84	98.09	98.47 (3)	98.39	77.01	86.39 (6)	7.25
DAD with timeseries	MadGAN	29.41	76.40	42.47 (7)	13.20	60.57	21.67 (7)	96.73	99.55	98.12 (6)	98.72	77.60	86.89 (2)	5.5
DAD with timeseries	BeatGAN	55.11	45.33	49.74 (5)	8.04	76.58	14.56 (12)	97.54	98.94	98.23 (5)	88.37	76.41	81.95 (12)	8.5
DAD with timeseries	OmniAnomaly	27.70	79.67	41.11 (9)	8.55	78.73	15.42 (11)	97.63	99.69	98.65 (2)	99.01	77.06	86.67 (4)	6.5
DAD with timeseries	MSCRED	61.26	59.11	60.17 (2.5)	55.80	34.32	42.50 (3)	97.31	99.43	98.36 (4)	98.43	77.69	86.84 (3)	3.125
text AD	CVDD	56.05	64.95	60.17 (2.5)	49.65	38.36	43.30 (2)	96.37	98.75	97.54 (9)	97.33	73.21	83.56 (10)	5.875
proposed	THOC	54.78	75.00	**63.31** (1)	61.50	36.34	**45.68** (1)	98.20	99.54	**98.86** (1)	98.08	79.94	**88.09** (1)	1.0

5. Ablation Study

5.1 Timeseries Representation and Hierarchical Structure

			$\mathcal{L}_{\text{orth}}$	$\mathcal{L}_{\text{TSS}}$	prec	recall	F1
flat	one hypersphere	RNN-top	×	✓	31.67	75.70	44.66
		USRL	×	×	59.49	27.10	37.24
	multiple hyperspheres	RNN-top	✓	✓	41.32	68.93	51.66
		USRL	✓	×	50.80	45.40	47.95
hierarchical	one hypersphere	THOC-variant	×	✓	53.27	60.98	56.86
	multiple hyperspheres	THOC	✓	✓	54.78	75.00	**63.31**



5. Ablation Study

5.2 Effectiveness of L_{orth} and L_{TSS}

$\mathcal{L}_{\text{orth}}$	$\mathcal{L}_{\text{TSS}}$	prec	recall	F1
×	×	52.22	24.77	33.60
✓	×	34.00	67.29	45.17
×	✓	42.08	57.71	48.67
✓	✓	54.78	75.00	**63.31**

감사합니다.