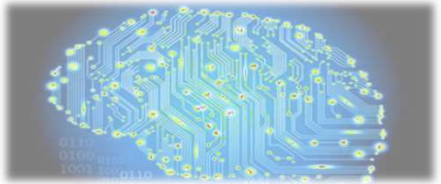
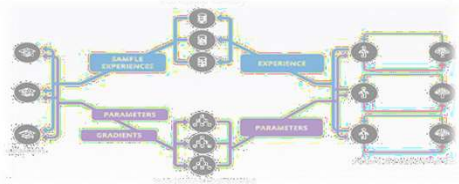
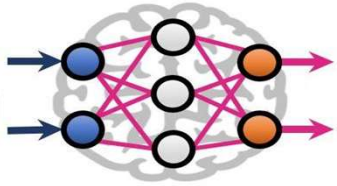
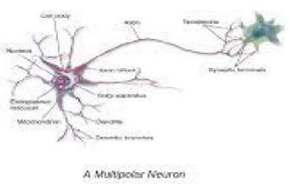
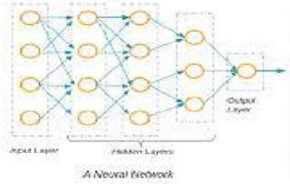







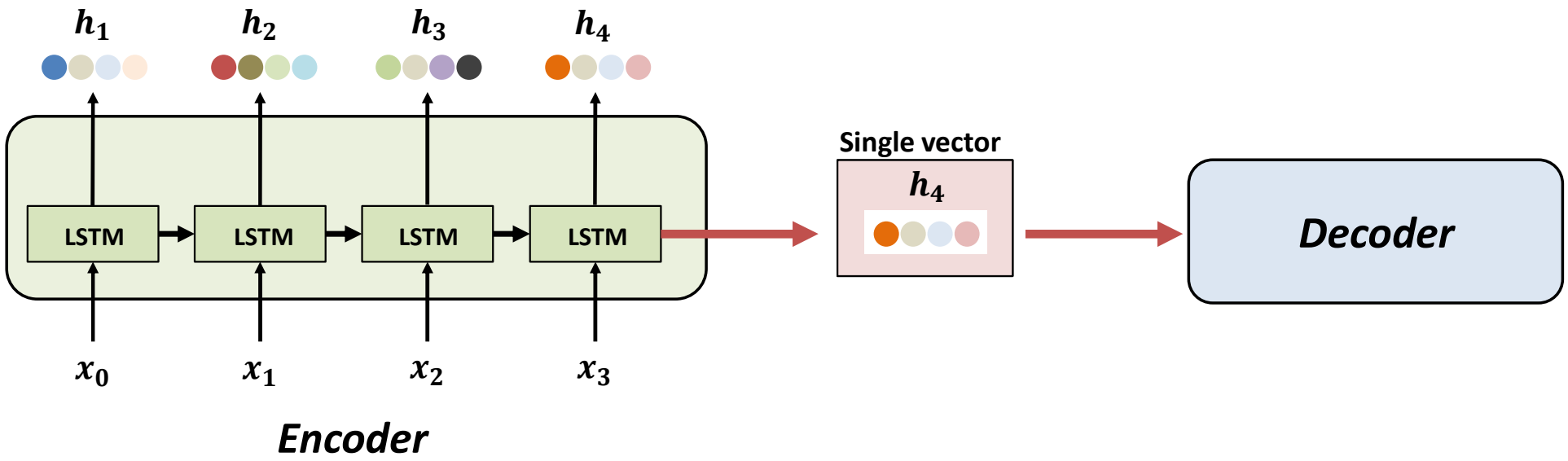
Comparison LSTM, GRU, Encoder-Decoder, CNN network and Attention

LSTM, GRU, Encoder-Decoder, CNN, Attention 모델 구현 및 실험

Geonil Yun,
rjsdlf45@gmail.com
2021. 07. 20.

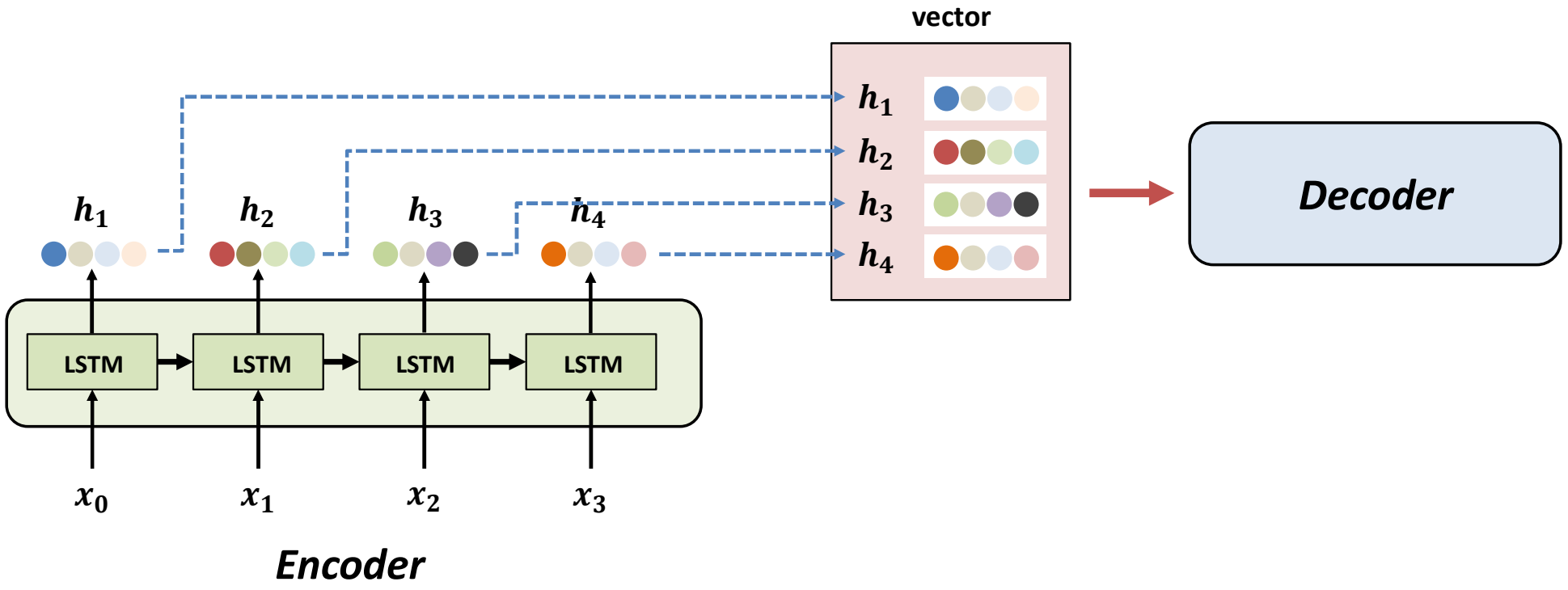
Introduction

- **Problem of Encoder-Decoder (Sequence to Sequence)**
  - output of Encoder is fixed-length vector (h_t)
  - compress time series data into fixed-length vector
  - not good at represent long time series data



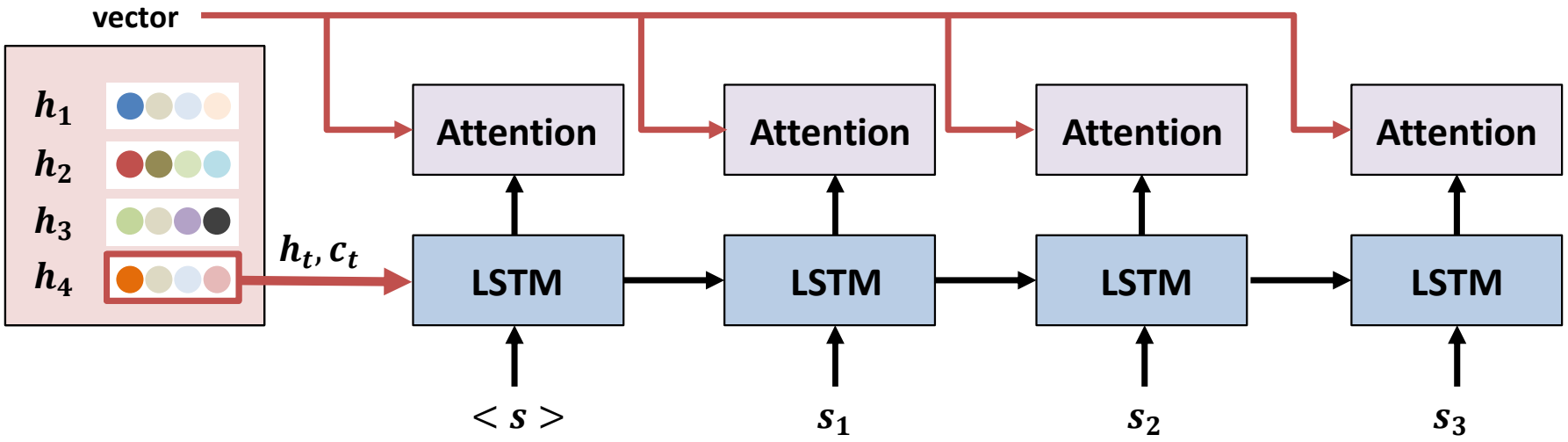
Proposed Method

- Improvements of Encoder-Decoder (Sequence to Sequence)
  - use all hidden states of Encoder in Decoder(h_1 … h_t)
  - good at represent each time step



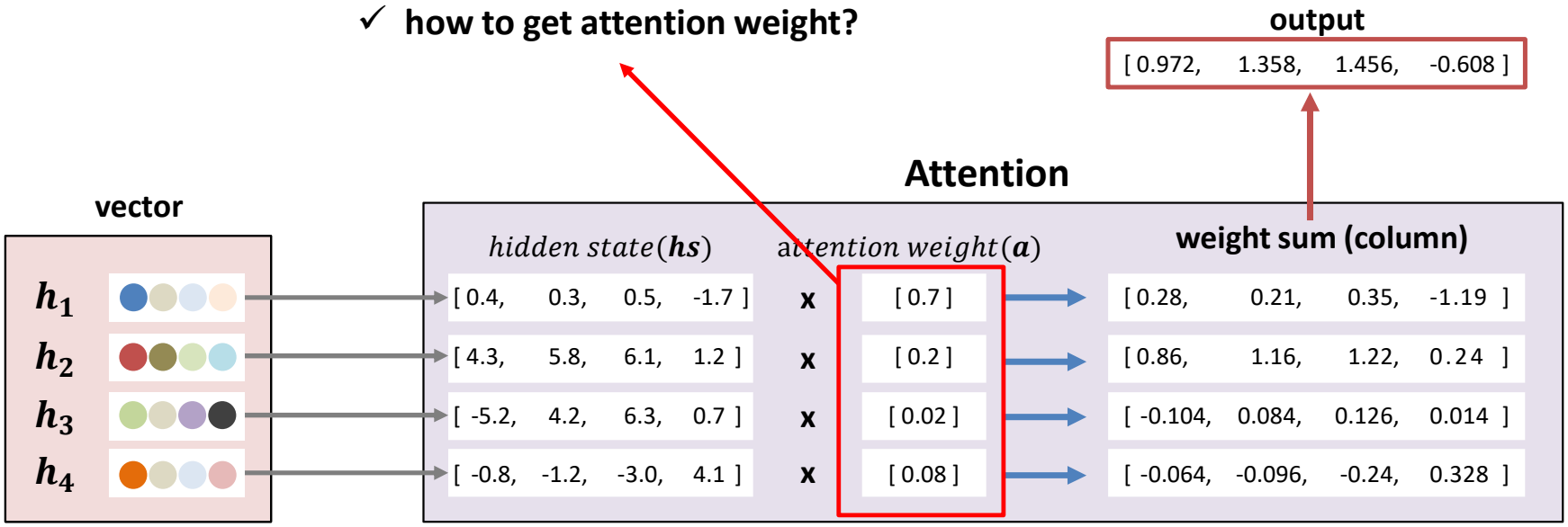
Proposed Method

- Attention
  - In Decoder, set hidden state and use all hidden states in next step
  - Attention to the hidden state that most similar to the current time step
  - Hidden state vectors($[h_1, \ \dots \ , \ h_t]$) are used every time step of Attention layer



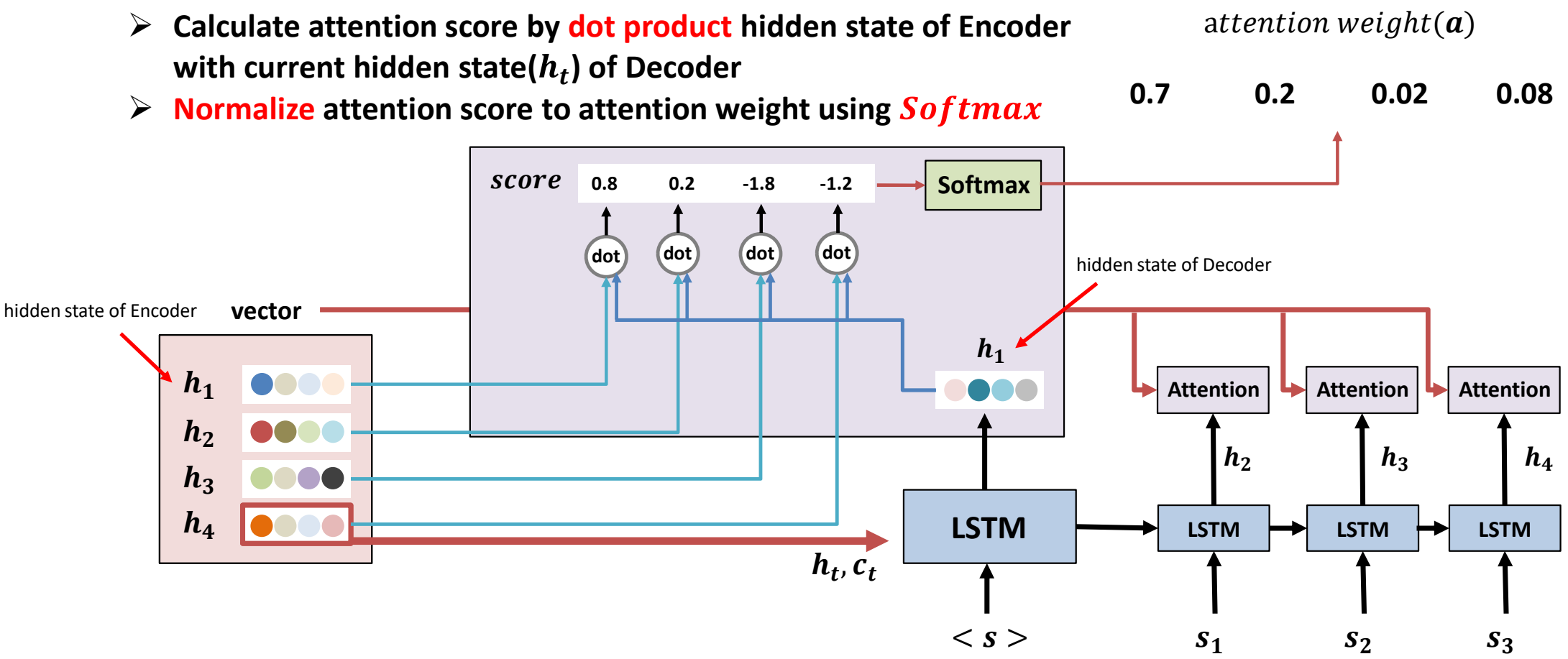
Proposed Method

- **Attention (weight sum)**
  - ➢ **Attention has attention weight(a) that can be learned**
  - ➢ **output is weight sum of hs x a**



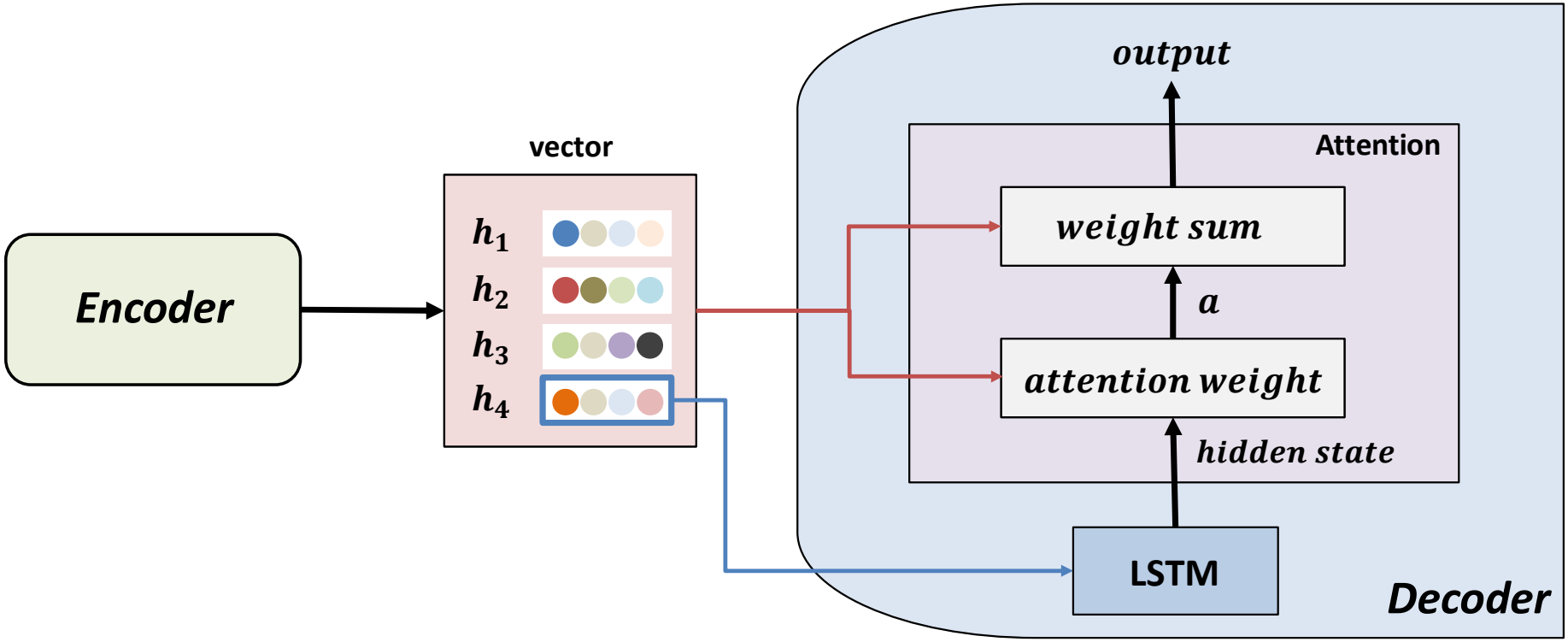
Proposed Method

- **Attention (attention weight)**
  - ➢ **Calculate attention score by dot product hidden state of Encoder with current hidden state(h_t) of Decoder**
  - ➢ **Normalize attention score to attention weight using $Softmax$**

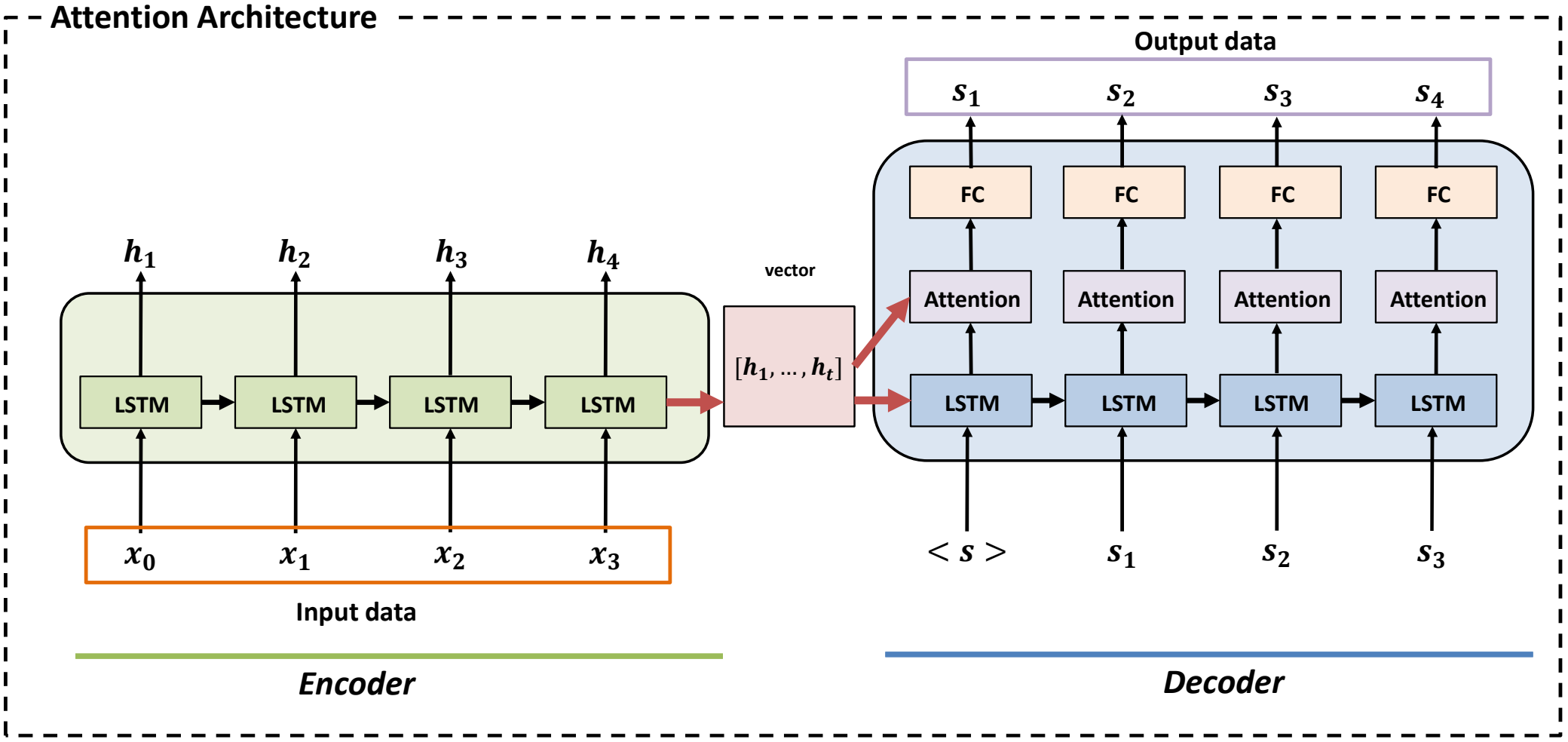


Proposed Method

- **Attention**
  - **Attention weight : use hidden state(h_t) and encoder vector**
  - **Weight sum : use encoder vector and attention weight(a)**

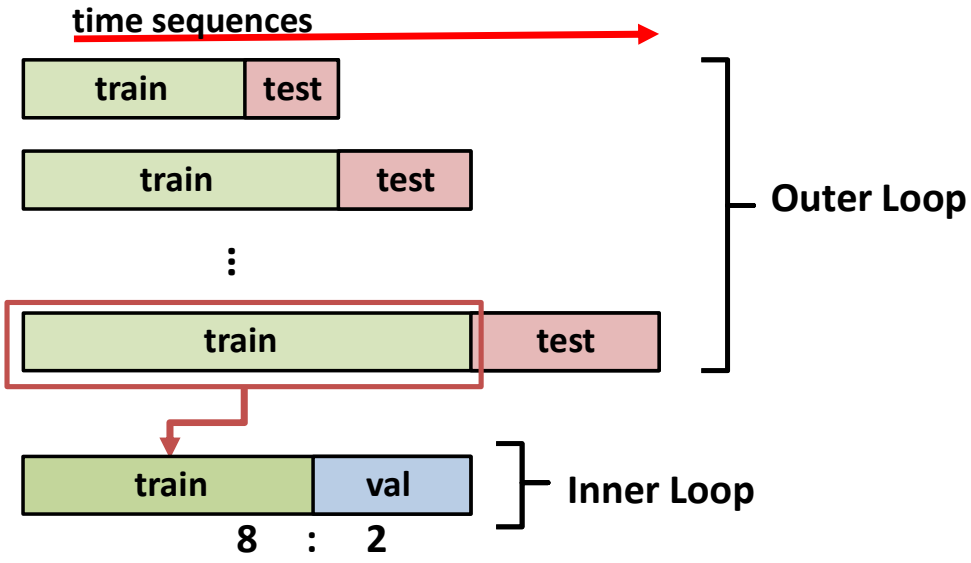


Proposed Method



Experimental setting

- **Nested Cross Validation for time series**



- ➢ **train past data & predict future data**
- ➢ **handling time series data in Nested CV**

- hyperparameter

Model	CNN (ShuffleNet)	RNN(LSTM, GRU)
fold num	10	
hidden size	64	
input length	24	
output length	1	
epoch	30	
batch size	32 or 64	
learning rate	**0.01, 0.001**	

- ➢ batch size 32 for **ISE, Chickenpox** dataset
- ➢ batch size 64 for others

Experimental setting

- **Statistics**

Dataset		Mean	Standard Deviation	Skewness	Kurtosis	Jarque Bera	ADF
Air Quality		1099.833	217.080	0.755	0.334	897.811	-9.823
Appliances Energy		97.694	102.524	3.386	13.664	191240.781	-21.616
Beijing Air Pollution		98.613	92.050	1.802	4.768	62162.743	-20.606
Hungarian Chickenpox		101.245	76.354	0.948	1.400	121.010	-6.933
Electric Consumption		1.483	1.194	1.141	1.732	3421.560	-5.360
Eye State		0.448	0.497	0.205	-1.957	2497.788	-4.803
Istanbul Stock Exchange	ISE	0.0016	0.016	-0.094	1.391	44.022	-22.746
	SP	0.0006	0.014	-0.090	3.068	211.013	-11.031
	DAX	0.0007	0.014	-0.106	2.104	99.894	-23.056
Metro Volume		3259.818	1986.860	-0.089	-1.309	3506.110	-28.016
Occupancy Detection		0.212	0.408	1.406	-0.020	2686.285	-3.430
SML 2010		20.493	3.312	-0.169	-0.105	21.792	-4.314

*ISE : Istanbul Stock Exchange, *SP : Standard & Poor's, *DAX : Germany

Comparison results

- Training time [s]

Data		CNN (ShuffleNet)	LSTM	GRU
Air Quality		85.945		
Appliances Energy		193.282		
Beijing Air Pollution		99.431		
Hungary Chickenpox		8.842		
Electric Consumption		100.391		
Eye State		154.765		
Istanbul Stock Exchange	ISE	9.133		
	SP	9.142		
	DAX	9.271		
Metro Interstate Traffic Volume		492.337		
Occupancy Detection		81.002		
SML 2010		41.084		

Comparison results

- RMSE (Root Mean Squared Error)

Data		CNN (ShuffleNet)	Standard Deviation	LSTM	Standard Deviation	GRU	Standard Deviation	LSTM Enc-Dec	Standard Deviation
Air Quality		147.489	33.383						
Appliances Energy		72.338	8.558						
Beijing Air Pollution		48.369	11.659						
Hungary Chickenpox		63.992	17.450						
Electric Consumption		0.848	0.357						
Eye State		0.059	0.074						
Istanbul Stock Exchange	ISE	0.028	0.007						
	SP	0.022	0.009						
	DAX	0.024	0.009						
Metro Interstate Traffic Volume		864.935	252.857						
Occupancy Detection		0.388	0.151						
SML 2010		2.512	1.128						

Comparison results

- MAE (Mean Absolute Error)

Data		CNN (ShuffleNet)	Standard Deviation	LSTM	Standard Deviation	GRU	Standard Deviation	LSTM Enc-Dec	Standard Deviation
Air Quality		112.998	25.412						
Appliances Energy		34.821	6.233						
Beijing Air Pollution		31.581	5.880						
Hungary Chickenpox		47.789	11.065						
Electric Consumption		0.623	0.294						
Eye State		0.469	0.068						
Istanbul Stock Exchange	ISE	0.022	0.006						
	SP	0.017	0.008						
	DAX	0.019	0.007						
Metro Interstate Traffic Volume		576.552	166.768						
Occupancy Detection		0.222	0.097						
SML 2010		2.321	1.146						

References

[1] : Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

[2] : Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Thank you

Appendix : Shuffle Net Architecture

Layer	Output size	KSize	Stride	Repeat	Output channels (g groups)				
					$g = 1$	$g = 2$	$g = 3$	$g = 4$	$g = 8$
Image	224×224				3	3	3	3	3
Conv1	112×112	3×3	2	1	24	24	24	24	24
MaxPool	56×56	3×3	2						
Stage2	28×28		2	1	144	200	240	272	384
	28×28		1	3	144	200	240	272	384
Stage3	14×14		2	1	288	400	480	544	768
	14×14		1	7	288	400	480	544	768
Stage4	7×7		2	1	576	800	960	1088	1536
	7×7		1	3	576	800	960	1088	1536
GlobalPool	1×1	7×7							
FC					1000	1000	1000	1000	1000
Complexity					143M	140M	137M	133M	137M

Appendix : Dataset

Dataset	references
Air Quality	https://archive.ics.uci.edu/ml/machine-learning-databases/00360/
Appliances Energy	https://archive.ics.uci.edu/ml/machine-learning-databases/00374/
Beijing Air Pollution	https://archive.ics.uci.edu/ml/machine-learning-databases/00381/
Hungarian Chickenpox	https://archive.ics.uci.edu/ml/machine-learning-databases/00580/
Electric Consumption	https://archive.ics.uci.edu/ml/machine-learning-databases/00235/
Eye State	https://archive.ics.uci.edu/ml/machine-learning-databases/00264/
Istanbul Stock Exchange	https://archive.ics.uci.edu/ml/machine-learning-databases/00247/
Metro Volume	https://archive.ics.uci.edu/ml/machine-learning-databases/00492/
Occupancy Detection	https://archive.ics.uci.edu/ml/machine-learning-databases/00357/
SML 2010	https://archive.ics.uci.edu/ml/machine-learning-databases/00274/