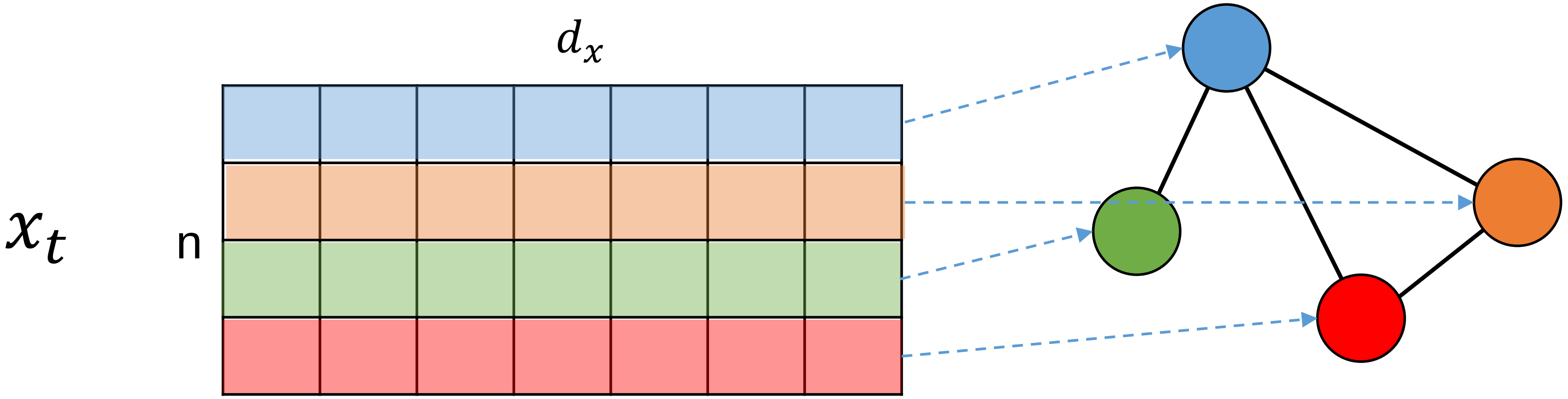
Structured Sequence Modeling With Graph Convolutional Recurrent Networks

Y.Seo, M.Defferrard, P.Vandergheynst, X.Bresson

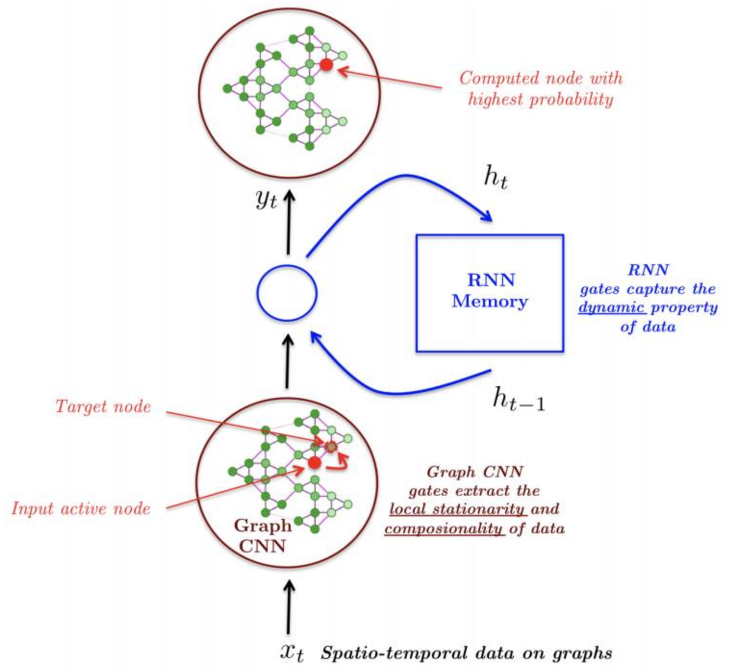
Presenter : Bumho Son

Objective of Research

- How to deal with structured sequence data?
- Spatio-temporal sequences
  - $\hat{x}_{t+1}, \dots, \hat{x}_{t+K} = argmax_{x_{t+1},\dots,x_{t+K}} P\left(x_{t+1}, \dots, x_{t+K} \middle| x_{t-J+1}, \dots, x_t\right)$
  - Consider data x_t as a graph signal : features are linked by pairwise relationship



Proposed Models : GCRN



- Graph Convolutional Recurrent Network (GCRN)
- Main idea
  - Use Graph CNN and RNN
  - Graph CNN : Identify spatial structures
  - RNN : Find dynamic patterns
- Model 1 : Stack a graph CNN on an LSTM

Graph CNN

$$x_t^{\text{CNN}} = \text{CNN}_{\mathcal{G}}(x_t)$$

LSTM

$$\begin{aligned}
i &= \sigma(W_{xi}x_t^{\text{CNN}} + W_{hi}h_{t-1} + w_{ci} \odot c_{t-1} + b_i), \\
f &= \sigma(W_{xf}x_t^{\text{CNN}} + W_{hf}h_{t-1} + w_{cf} \odot c_{t-1} + b_f), \\
c_t &= f_t \odot c_{t-1} + i_t \odot \tanh(W_{xc}x_t^{\text{CNN}} + W_{hc}h_{t-1} + b_c), \\
o &= \sigma(W_{xo}x_t^{\text{CNN}} + W_{ho}h_{t-1} + w_{co} \odot c_t + b_o), \\
h_t &= o \odot \tanh(c_t).
\end{aligned}$$

Graph filtering operation

$$y = g_\theta *_{\mathcal{G}} x = g_\theta(L)x = \sum_{k=0}^{K-1} \theta_k T_k(\tilde{L})x,$$

θ : Chebyshev coefficients

$T_k(\tilde{L})$: Chebyshev polynomial of order k

Proposed Models : GCRN

- Model 2 : Generalize the convLSTM model to graphs

$$
\begin{aligned}
i &= \sigma(\boxed{W_{xi} *_{\mathcal{G}} x_t} + W_{hi} *_{\mathcal{G}} h_{t-1} + w_{ci} \odot c_{t-1} + b_i),\\
f &= \sigma(W_{xf} *_{\mathcal{G}} x_t + W_{hf} *_{\mathcal{G}} h_{t-1} + w_{cf} \odot c_{t-1} + b_f),\\
c_t &= f_t \odot c_{t-1} + i_t \odot \tanh(W_{xc} *_{\mathcal{G}} x_t + W_{hc} *_{\mathcal{G}} h_{t-1} + b_c),\\
o &= \sigma(W_{xo} *_{\mathcal{G}} x_t + W_{ho} *_{\mathcal{G}} h_{t-1} + w_{co} \odot c_t + b_o),\\
h_t &= o \odot \tanh(c_t).
\end{aligned}
$$

$W_{xi}x_t$ Classic LSTM : matrix multiplication by dense matrix W

$\downarrow$

$W_{xi} * x_t$ convLSTM : Replace multiplication by 2D convolution($*$) by a set of kernels (Shi et al.(2015))

$\downarrow$

$W_{xi} *_{\mathcal{G}} x_t$ GCRN : Replace 2D convolution by the graph convolution($*_{\mathcal{G}}$)